The Astronomical Application of Infrared Array Detectors

Mark John McCaughrean

University of Edinburgh

Doctor of Philosophy

1987

This thesis has been composed by me, and consists of my own work, except where specifically indicated in the text.

December 1987

Contents

List of Figures

List of Tables

Acknowledgments

After spending five years at the Royal Observatory Edinburgh, I owe much to its staff, both here in Scotland, and in Hawaii. I thank my supervisor, Ian McLean, for involving me in the new technology of infrared arrays, and for his guidance, support, and advice, even when the pressure of running the Infrared Array Project made his own time precious. Mike Smyth of the University of Edinburgh Astronomy Department, my other supervisor, has supported me also, from my undergraduate studies until the present time. As the Infrared Array Project called upon the resources of almost every unit, thanks are due to almost every member of staff. For the design and construction of an excellent infrared camera system, IRCAM, particular thanks to the staff and managers of the Technology Unit and the Workshops, including Tim Chuter, Terry Lee, Donald Pettie, Ron Beetles, John Clark, Marjorie Fretwell, Frank Holmes, Davie Laird, Magnus Patterson, Ian Smith, and Bharat Vyas. Amongst the other ROE staff, I would like to thank Richard Wade, Matt Mountain, Hans Zinnecker, and Maureen McLean of the UKIRT Unit; Denis Kelly and John Barrow of the Image and Data Processing Unit; Brian Hadley, Bill Roberton, and Mairi Martin of the Photolabs; Sue Tritton of the Plate Library; Peter Brand, Liz Gibson, and Susan Hooper of the University Astronomy Department; the security staff for tolerating my peculiar hours; and the Director of the ROE, Malcolm Longair, for his tolerance and considerable financial support during my long association with the Infrared Array Project.

In Hawaii, I would like to thank all the staff of the UKIRT Unit, including Malcolm Smith, David Beattie, Tom Geballe, Anna Lucas, Yolanda Boyce, Clive Campbell, Jay Tsutsumi, Kent Tsutsui, and particularly the night assistants, Dolores Walther and Joel Aycock, whose superb skills at the 'sharp end' made the whole effort worthwhile.

Outside the ROE organisation, my thanks are due to Alan Hoffman of SBRC for advice and useful discussions on the SBRC array; Craig McCreight and Mike Werner of NASA Ames Research Center, and Dan Gezari of NASA Goddard Space Flight Center for financial support and opportunities to give seminars; the Lunar and Planetary Laboratories in Tucson, Arizona for use of the 1.54m telescope on Mount Bigelow; the Jesuits of the Vatican Observatory, George Coyne, Rich Boyle, Charlie Polzer, Chris Corbally, and Philippe Eenens, for their hospitality at Kino House and valued friendship.

My fellow postgraduates have been a source of friendship, advice, and useful discussion. I would particularly like to thank John Rayner, Mark Bird, Mike Burton, and Richard Prestage. Before ending my astronomical acknowledgements, I would like to thank Colin Aspin, for his friendship, hospitality, advice, help, and many hours of companionship during long observing trips in the United States and the Canaries, in addition to his considerable contributions to the success of the Infrared Array Project.

This thesis was typeset using TEX and LATEX, and I am grateful to Donald Knuth and Leslie Lamport for their implementations of these systems, and for making them publicly available.

I acknowledge the financial support of the Science and Engineering Research Council throughout the majority my studies, at first under a Research Studentship, and later via the ROE as a research assistant.

Finally, many people outside of astronomy are to be thanked for their compassion and understanding, but particularly Claire for her faith in me, and my mother and brothers for their unending support and love. Lastly, and most dearly, I thank Sande, for everything.

Abstract

Major innovations in semiconductor technology have led to the recent availability to the astronomical community of arrays comprising several thousands of detectors, each simultaneously sensitive to near infrared radiation. The application of these infrared array detectors to astronomy is revolutionising the way we see the sky at these wavelengths.

Whilst similar in concept to digital imaging arrays as used by optical astronomers, there are nevertheless many areas in which the design, implementation, and optimisation of imaging arrays used in ground based near infrared astronomy differ significantly from those applied to their optical counterparts. This thesis examines the requirements for detector arrays used for astronomical near infrared imaging, and in particular examines the operation and optimisation of the recently available SBRC 62 x 58 pixel indium antimonide hybrid focal plane array. I develop a model of this array that allows us to simulate its behaviour across a wide range of observational configurations, and how its inherently non-linear response to photon illumination may be characterised and compensated for. I also develop a model that allows us to predict the sensitivity of this detector array when used in the new near infrared imaging camera (IRCAM), recently commissioned at the United Kingdom 3.8 metre Infrared Telescope on Mauna Kea, Hawaii.

Finally, I present novel infrared images of the Orion Nebula as obtained with IRCAM during the instrument commissioning. I examine the techniques used to make the observations, and to reduce the many individual frames into one single continuous mosaic. I compare the new images with previous data, and discuss the nature of several new infrared sources seen in our images, particularly with regards young stars embedded in the gas and dust of the region. These large scale, seeing limited near infrared images clearly demonstrate the large step forward taken by infrared astronomy with the introduction of imaging arrays.

Chapter 1

Introduction

Soon as the sound had ceased whose thunder filled
The abysses of the sky and the wide earth,
There was a change ... the impalpable thin air
And the all-circling sunlight were transformed,
As if the sense of love, dissolved in them,
Had folded itself round the spherèd world.
My vision then grew clear, and I could see
Into the mysteries of the universe.

Spirit of the Hour
"Prometheus Unbound"
Percy Bysshe Shelley

In 1978, a new generation of 2–4 metre class dedicated infrared telescopes were nearing completion, including the United Kingdom 3.8 m Infrared Telescope, the NASA 3 m Infrared Telescope Facility, and the 2.3 m Wyoming Infrared Observatory. At the same time, infrared astronomical instrumentation was approaching the limits of its current line of development. Single channel photometers based on single element detectors of highly optimised semiconductor materials had almost reached their maximum theoretical sensitivity.

In the following decade, no major new infrared telescopes have been built, and instrumentation for the infrared has remained dependent on the single element detector. In the same intervening period, the infrared has become recognised as a critical part

of the electromagnetic spectrum, particularly for studies related to the star formation process. Accordingly, infrared astronomy has become a major branch of astrophysical research, and the number of papers published annually in the field continues to rise steeply.

Despite the burgeoning demand, technological advances have been slow, and only slight gains in overall sensitivity have been achieved. In concluding their review of infrared instrumentation in 1978 however, Soifer and Pipher noted that in years to come, one technological advance would "...drastically affect infrared astronomy ... [and] should provide a major increment in the sensitivity of astronomical observations." [131]. The technology is that of infrared arrays, and today they are finally a reality.

After years of experimentation with various off-the-shelf detector materials and array technologies, infrared instrument builders now have access to several highly suitable types of detector array, and in the past year, common-user infrared cameras based on these arrays have become available at three major observatories.

One such detector array is the SBRC 62×58 pixel array of indium antimonide (InSb) photo-diodes, read-out by a silicon switched FET or direct readout (DRO) multiplexer. One such camera is IRCAM, a multi-purpose 1–5 μm imaging system for the United Kingdom 3.8 Infrared Telescope, on Mauna Kea, Hawaii. This camera system was developed at the Royal Observatory Edinburgh under project scientist Dr. Ian McLean, as part of the ROE Infrared Array Project. During the design, development, construction, and commissioning of this system, I have been a member of the Infrared Array Project team, and the work presented in this thesis documents my contributions to that project.

In chapter 2, we examine the technical background to the new infrared array technology, and why it has been so long in coming, compared to the solid state detectors such as CCDs, as used by optical astronomers. We briefly review the basic principles of infrared detection, and the status of astronomical infrared instrumentation prior to the arrival of arrays. We examine the advantages and disadvantages of infrared arrays, and some of the parameters that are particularly vital to their use in astronomy.

In chapter 3, we examine in detail the detector physics behind the SBRC 62×58 InSb + DRO array, with particular emphasis on deriving a detector model that can be used to

simulate the operation of the array. We use this model to examine the behaviour of the array in a range of likely astronomical observing conditions, with particular emphasis on the non-linear photon detection inherent in this array. We discuss the magnitude of the non-linearity in these likely regimes, and also how the non-linearity may be corrected. We also develop a model of the noise characteristics of the array, and apply it to actual data, in order to demonstrate the degree to which the array is limited by background shot noise, read noise, and residual spatial nonuniformities.

In chapter 4, we discuss the development of SIRCAM, a software package designed to simulate near infrared imaging systems. The package was designed in order to anticipate the problems that may arise in these systems. Particular attention is paid to the very large range in sky background flux measured across the 1–5 μm region, and the thermal and non-thermal sources that give rise to it. We discuss the application of this package to IRCAM, and using the information discussed in this chapter, and the detector model derived in chapter 3, we make sensitivity predictions across the wide range of observational configurations possible in IRCAM, from seeing limited imaging of point sources, to low resolution imaging of low surface brightness extended emission. We also show some simulated images, in order to prepare the reader for the kind of data that might be expected from this first generation infrared imaging system. We finally examine some problems associated with heavily undersampled imaging, in particular photometric errors and aliasing.

In chapter 5, we present data obtained during the early part of the commissioning period of IRCAM on the UKIRT, in particular a $5.3' \times 4.7'$ image of the Orion Nebula (M42), made at 2.2 μm with seeing limited resolution. This image is a mosaic of 126 individual IRCAM frames, and in the first half of the chapter we discuss the observing strategy and data reduction techniques used to obtain and combine these data. Examples are given of the data at various stages of the processing, and comparisons of several possible reduction techniques are briefly made. The second half of chapter 5 is devoted to an examination of these data from an astronomical point of view. We compare it with previous maps and images of the region in order to assess the new imaging technique. Using our new 2.2 μm image, and limited imaging of the same region at 1.6 μm and 3.6 μm, we then discuss several new features seen in the region in the infrared. We discuss a cluster of infrared point sources that are located along the line of sight towards a secondary condensation in the molecular cloud OMC-1, and discuss the possibility of

their being embedded in this condensation, providing its luminosity. We also discuss the discovery of new infrared point sources near to known Herbig-Haro-like objects, infrared identifications of a dense cluster of compact radio sources, and the 2.2 μm frequency function of 440 members of the extremely dense cluster of low mass stars that surrounds the well-known Trapezium OB association.

Appendix A lists the accurate positions of sources discussed in chapter 5, as derived from astrometric measurements from within the 2.2 μm image, which we have keyed into the fundamental astrometric frame of reference via specially commissioned far red Schmidt plates of the region. The positions are given at both B1950.0 and J2000.0 equinoxes, and are accurate to $\pm 1''$ RMS.

Appendix B contains a brief summary of the image processing package Rapi2d, which was used to reduce the data presented in chapter 5, and was written by this author as part of his work on the Infrared Array Project.

Appendix C contains copies of papers involving the author published during his postgraduate work.

1.1 IRCAM

Before proceeding into the body of this thesis, we shall briefly discuss IRCAM, the new common-user 1–5 μm infrared array camera, currently in operation on the United Kingdom 3.8 m Infrared Telescope, on Mauna Kea, Hawaii, and built at the Royal Observatory Edinburgh by the Infrared Array Project team. A copy of a paper more fully describing the instrument is to be found at the back of this thesis [99], and further detailed information is to be found elsewhere [141,120]. However, as the design and implementation of the camera system are not the subject of this thesis, we shall only briefly discuss it here, in order to introduce the specific features of the camera, which are relevant to this work, and to which we shall refer repeatedly throughout the thesis.

At the heart of IRCAM is an SBRC 62 × 58 InSb + DRO infrared detector array, and this array is the subject of a detailed analysis in chapter 3 of this thesis. The detector array is housed inside a cryogenically cooled chamber known as the camera body. Also inside the camera body are two filter wheels, containing a wide range of

infrared interference filters. The range of filters includes the standard astronomical near infrared broad band set (J, H, K, L′, M), and several narrow band interference filters, designed for imaging spectroscopy in the lines of, for example, the v=1–0 S(1) line of molecular hydrogen at 2.122 μm, the Brγ line of ionised hydrogen at 2.166 μm, and the Brα line of ionised hydrogen at 4.05 μm. Also included is a special narrow band filter with a central wavelength of 3.6 μm, and a bandwidth of 1.8%. This latter filter is used for imaging at a wavelength where the thermal background flux seen through the conventional broad band filter saturates the detector array before it can be read out.

The other main feature of IRCAM and of importance to this thesis, in particular to the SIRCAM modelling work in chapter 4, is the option of different pixel scales on the sky. Three different pixel scales are available by interchanging ZnSe lens modules within the camera body — the scale sizes are 0.6, 1.2, and 2.4″/pixel. The former is used for high spatial resolution imaging, and point source photometry. The latter two scale sizes are designed to provide increased sensitivity to very low surface brightness extended emission. The 0.6″/pixel mode almost fully samples the mean seeing on Mauna Kea, but the other two do not, and this undersampling has serious consequences for accurate photometry and unambiguous detection.

Finally, the external optics. The camera body is supported on an optical table, which is in turn connected to the UKIRT Cassegrain focus instrument support unit (ISU2). The beam of converging infrared radiation coming down from the secondary mirror is intercepted by a dichroic mirror in ISU2, which deflects it through 90°. The beam is diverging again by the time it reaches a warm gold coated flat mirror on the IRCAM optical table. This folds the beam, before the beam proceeds on to a warm gold collimator. This collimates the beam prior to entering the cryogenic camera body in order that warm Fabry-Pérot étalons may be placed immediately in front of the camera body, to provide higher spectral resolution than can be obtained with the narrow band filters alone.

All the other components of IRCAM are essentially transparent, in that they do their job without affecting the astronomical performance of the camera. We shall not consider them further in this thesis, except to acknowledge the enormous effort that went into the design and implementation of the whole camera system. The people involved are acknowledged by name elsewhere in this thesis.

Chapter 2

Infrared array technology review

In this chapter, we shall briefly examine the background to the emergence of infrared array technology. We shall discuss the reasons why photon detection in the infrared is harder than in the optical, and why solid state imaging arrays for the infrared have taken so much longer to become available to astronomers. We shall discuss the considerable advantages that infrared arrays bring to astronomy, and the one or two minor disadvantages. We shall examine the basic operating criteria of infrared arrays, and the parameters that are most important for their application to ground-based astronomy. We shall discuss some of the issues that arise when arrays of infrared detectors are used as opposed to the conventional single element detector, with regards sky subtraction, DC integration, and pixel-to-pixel non-uniformity.

Finally, we discuss the techniques used to manufacture infrared arrays, noting the various advantages and disadvantages associated with each. Before proceeding to examine the SBRC InSb + DRO array in detail in the following chapter, we briefly mention the status of other arrays available for use in the near infrared, and possible future developments of infrared array detectors for astronomical applications.

2.1 Detection of infrared radiation

The infrared region of the electromagnetic spectrum straddles the boundary between regimes towards shorter wavelengths where light is generally treated in terms of photons,

and longer wavelengths, where the wave properties of light are considered most important. Infrared detection techniques cover the range the whole range of possibilities. In heterodyne systems for the sub-millimetre for example, wave interaction techniques are employed, as at radio wavelengths. In the mid and far infrared, bolometers rely on thermal effects, where the energy of the incoming radiation heats the detector, and creates a measurable change in its electrical properties. At the shorter infrared wavelengths, between 1 and 25 μm, the effects on a semiconductor material of individual photons are measured.

The many possible ways of detecting infrared radiation are reviewed in detail in many standard texts [79,158,159,77]. In addition, Gillett *et al.* [51] gave an excellent review of the detectors that were in use in infrared astronomy towards the end of the 1970s. We shall briefly discuss the basic principles of photon detectors only, immediately concentrating on the internal photon effects, in which incoming photons free charge carriers within a semiconductor material, and which are employed by virtually all astronomical detectors used at wavelengths shortward of 25 μm.

2.1.1 Intrinsic and extrinsic photon detection

Photon detectors are made from semiconductor materials, where an incoming photon with sufficient energy will create an electron-hole pair, a free electron, or a free hole, depending on the mechanism employed. In intrinsic semiconductors, a photon with sufficient energy will create an electron-hole pair, with the electron entering the conduction band, and the hole entering the valence band. This is known as the intrinsic effect. Extrinsic semiconductors are doped with impurities, and an incoming photon with insufficient energy to excite an electron-hole pair in the semiconductor itself may still be energetic enough to produce excitation at an impurity centre. Either a free electron-bound hole, or bound electron-free hole may result. This is known as the extrinsic effect. The intrinsic and extrinsic effects are illustrated in figures 2.1 and 2.2 respectively.

Probably the most important parameter when choosing a semiconductor material for a detector is the cut-off wavelength λ_{co}, i.e. the longest wavelength of illumination to which the material is sensitive. In the case of intrinsic materials, this is determined

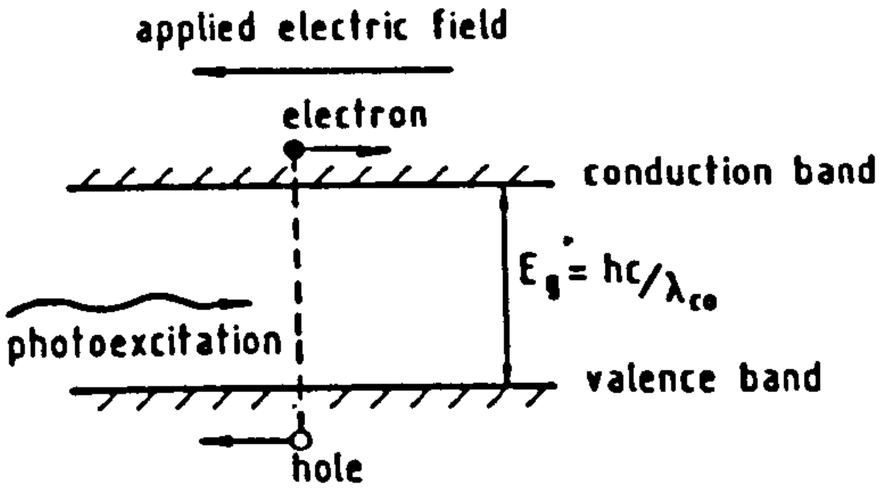


Figure 2.1: Intrinsic effect

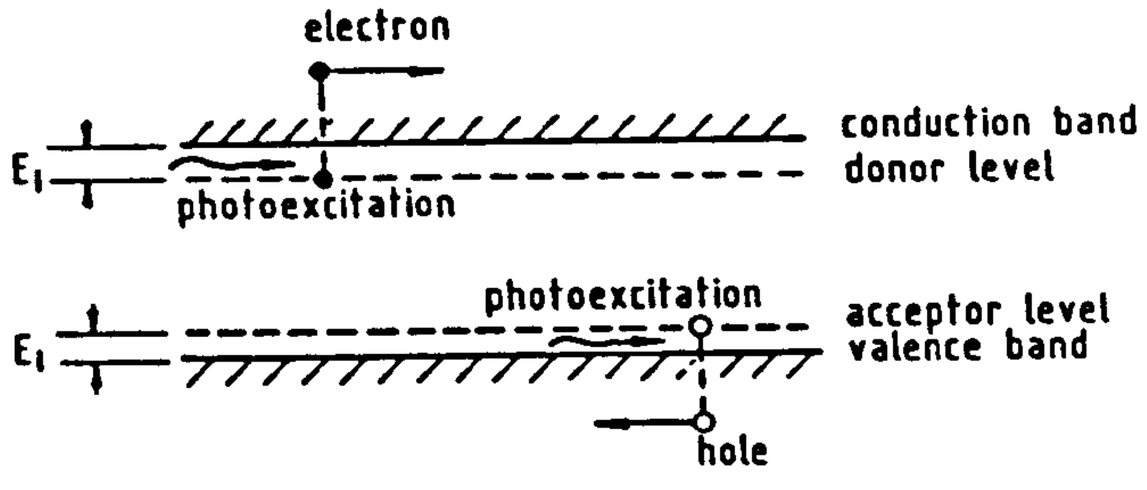


Figure 2.2: Extrinsic effect

by the semiconductor bandgap energy E_g, with the simple relation that $\lambda_{co} = hc/E_g$. This is frequently written in the following units :

$$\lambda_{co} \quad (\mu\text{m}) \qquad = \frac{1.24}{E_g} \quad (\text{eV}) \tag{2.1}$$

For example, indium antimonide has a bandgap energy of 0.23 eV at 77 K, and therefore has a cut-off wavelength at 5.4 μm. We shall see later in this thesis that the bandgap energy is temperature dependent, and thus that the cut-off wavelength changes with detector temperature.

For extrinsic materials, the cut-off wavelength is similarly defined by the ionisation energy E_i of the impurity material in the host semiconductor. For example, arsenic doped silicon has an ionisation energy of 0.0537 eV, and the corresponding wavelength cut-off can be calculated from above by substituting E_i for E_g, giving $\lambda_{co} = 23.1\,\mu$m.

Tables 2.1 and 2.2 give the bandgap energy and wavelength cut-off for a variety of intrinsic and extrinsic semiconductor materials respectively, with cut-off wavelengths ranging from the optical to the far infrared.

The narrow bandgap necessary to detect infrared photons implies that these materials must be cooled to much lower temperatures than semiconductors sensitive to optical radiation, in order to reduce thermally generated dark current. Typically, operating temperatures below the boiling point of liquid nitrogen at 77 K are necessary for 1–5 μm materials, and a liquid helium temperature of 4 K is required for longer wavelength detectors.

Material	T (K)	E_g (eV)	λ_{co} (μm)
CdS	295	2.40	0.52
CdSe	295	1.80	0.69
CdTe	295	1.50	0.83
GaP	295	2.24	0.55
GaAs	295	1.35	0.92
Si	295	1.12	1.11
Ge	295	0.67	1.85
PbS	295	0.42	2.95
PbSe	195	0.23	5.40
InAs	195	0.39	3.18
InSb	77	0.23	5.40
$Pb_{0.2}Sn_{0.8}Te$	77	0.10	12.4
$Hg_{0.8}Cd_{0.2}Te$	77	0.10	12.4

Table 2.1: Intrinsic semiconductors

Material : Impurity	E_i (eV)	λ_{co} (μm)
Ge : Au	0.15	8.27
Ge : Hg	0.09	13.8
Ge : Cd	0.06	20.7
Ge : Cu	0.041	30.2
Ge : Zn	0.033	37.6
Ge : B	0.0104	119.2
Si : In	0.155	8.00
Si : Ga	0.0723	17.1
Si : Bi	0.0706	17.6
Si : Al	0.0685	18.1
Si : As	0.0537	23.1
Si : P	0.045	27.6
Si : B	0.0439	28.2
Si : Sb	0.043	28.8

Table 2.2: Extrinsic semiconductors

It is worth noting that as well as being the prime material for sensing optical radiation, silicon is used for virtually all other semiconductor devices, and thus it is a very mature technology. In contrast, the more exotic materials required for the detection of infrared photons have few applications outside of infrared detection, and thus their technologies are considerably less developed.

Historically, lead sulphide (PbS) was the first detector used widely for infrared astronomy, but it was later superseded by indium antimonide, still the most widespread material for the near infrared [74]. In the 1960s and 1970s, mercury cadmium telluride (HgCdTe) became the preferred detector for many non-astronomical applications [70]. It is now a strong competitor for astronomical detectors also. It has the quality that the bandgap energy may be tuned, and therefore the effective wavelength cut-off is altered. This is done by altering the relative molar contributions of mercury telluride and cadmium telluride. The example listed in table 2.1 was a longer cut-off material — a 2.5 μm cut-off is achieved with $Hg_{0.554}Cd_{0.446}Te$ material [124].

2.1.2 Photoconductivity

By placing an external electric field across the semiconductor detector, the charge carriers liberated by the incoming photons may be separated before they have time to recombine. This creates a measurable photo-signal, detected either as a change in voltage across an external load resistor, or as a change in current across the detector. A schematic diagram of a typical photoconductive geometry is shown in figure 2.3. Both intrinsic and extrinsic semiconductor materials are used to detect radiation through the photoconductive effect.

2.1.3 Photovoltaic effect

Unlike the photoconductive effect, the photovoltaic effect does not use an externally applied field to separate the charge carriers generated by photon excitation. An internally created field provides the separation, usually in the form of a semiconductor p-n junction. Diffused into the material via standard techniques, the p-n junction creates a built-in electric field. Incoming photons may create electron-hole pairs, either in the junction itself, or in the bulk material. If the electron-hole pair sees the influence

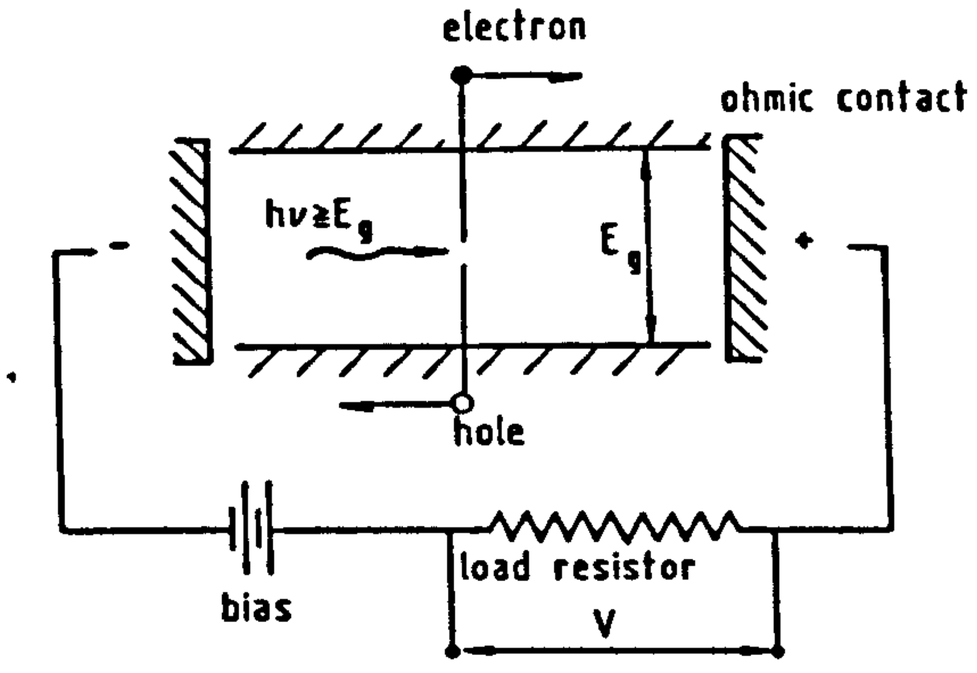


Figure 2.3: Photoconductive effect

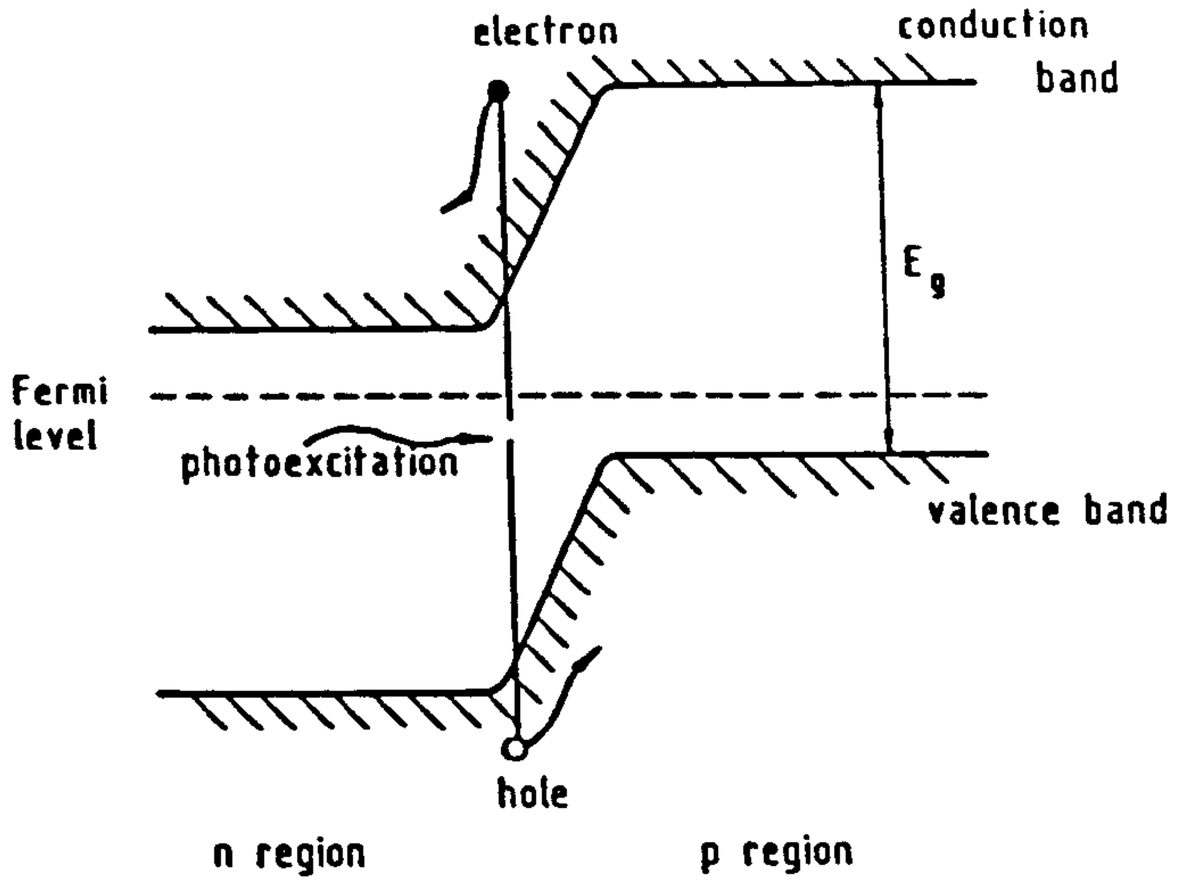


Figure 2.4: Photovoltaic effect

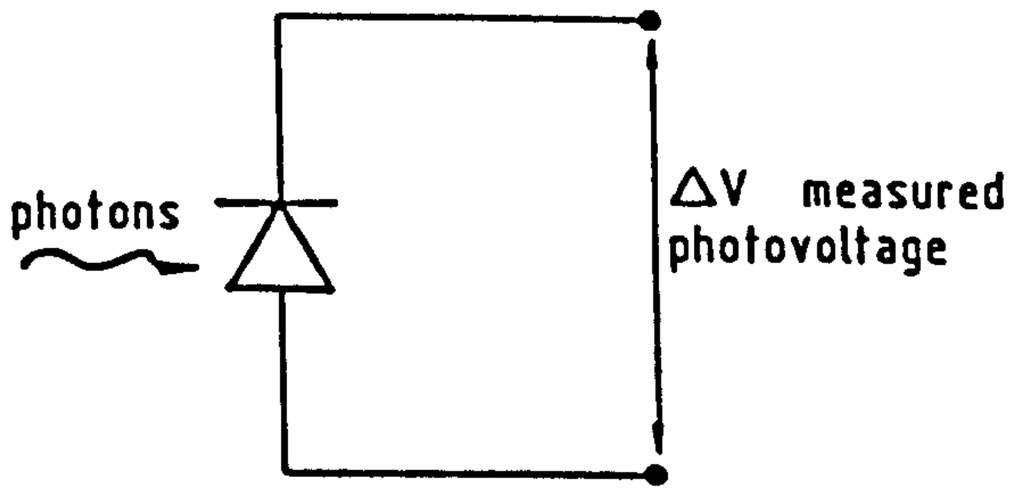


Figure 2.5: Open circuit

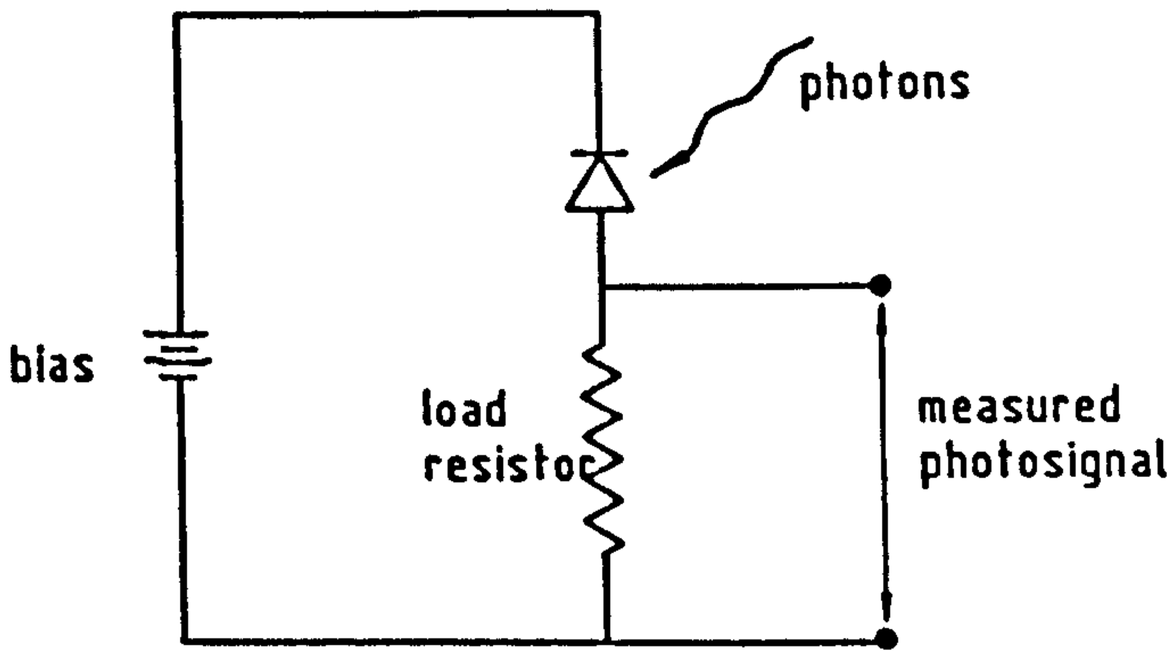


Figure 2.6: Reverse bias

of the p-n junction electric field before recombining, the field will separate the charge carriers. This is illustrated in figure 2.4.

With a constant photon flux, and with no external circuit placed across the detector, the detector will settle at a forward voltage known as the open circuit voltage, and this may be measured to infer the incoming photon flux, as shown in figure 2.5. This operation with no external circuit is the true photovoltaic effect. However, if a reverse bias is placed across the junction, a current will flow in the external circuit due to the detection of photons. This current may be detected as a change in voltage across an external load resistor, as shown in figure 2.6. In this configuration, the circuit is said to be operating in reverse bias photoconductive mode, although this is not to be confused

with the straightforward photoconductive effect discussed above.

Virtually all astronomical infrared single element detector systems operate in the reverse bias mode. In order to minimise the effects of excess noise, including $1/f$ components, the detector is held near zero bias, and the output photo-current flowing in the external circuit creates a voltage across a feedback resistor [57]. Careful electronics and amplifier design are required to maintain the detector near zero bias, but when this has been achieved, infrared detectors can be operated close to the fundamental limit where Johnson noise limits operation [57]. An additional benefit of holding the detector at zero bias, and measuring the resulting photo-current, is that the displacement of the *I-V* curve down the current axis is linearly proportional to the photon flux, whereas the open circuit photo-voltage is not [77, p. 16].

By applying these techniques to high quality InSb photodiodes, Hall *et al.* [57] demonstrated a trans-impedance amplifier (TIA) design that allowed Johnson noise limited performance. Single element detector systems using this approach are still the norm for infrared photometers.

Improvements in the TIA type design have been limited since its development. In the 1–2.5μm region, the photon background may be relatively low, and very high detector impedances ($\geq 10^{11}\Omega$) are required to ensure the system is limited by the background shot noise, and not the detector Johnson noise. Then the feedback resistor used in the TIA circuit must be at least three times larger than the detector resistance, in order that its Johnson noise does not dominate [74]. As detectors have improved, their resistances have increased, and it is difficult to find commensurately larger feedback resistors that maintain voltage linearity, critical for astronomical applications. Also, high photon fluxes can saturate the very large feedback resistor, and although this may be overcome by switching feedback resistors depending on the photo-current (e.g. the UKT9 photometer on UKIRT), the offset between the two resistors then must be calibrated.

2.1.4 AC coupled detection

Single channel detectors can only measure one point on the sky at a time. Because the infrared sky is bright and variable, and because the sources of interest are generally

much fainter than the sky, some method of rapid sky referencing is required to ensure that the sky background is adequately subtracted from the source signal. The chosen technique is that of 'chopping'. The infrared beam is rapidly switched between the source position and a nearby reference position, by use of a small focal plane mirror or a wobbling secondary. This chopping between source and sky takes place at about 3–10 Hz in the near infrared, and the output signals are phased with a reference signal from the chopper, and only the AC component is amplified, representing the source flux alone. In order to compensate for any asymmetries in the optics or linear gradients in the sky background, the source and sky beams are switched periodically, in a process known as 'nodding'.

The chopping technique allows for accurate cancellation of the background, but also reduces sensitivity to extended source emission where the chop throw is less than the scale length of the extended emission [86]. Indeed, this very feature may be used in complex regions to search for point sources, with small chop throws used to eliminate nebulous background [90]. However, the actual chop frequencies used are high and largely historically determined by the nature of early detectors, which were unstable. The chop frequency was chosen as a compromise between the settling time required due to internal detector RC time constants, and the desire to eliminate $1/f$ noise [14]. In the near infrared particularly, excess sky noise does not degrade performance until much lower chopping frequencies are used [9].

2.1.5 DC coupled detection

2.1.5.1 The AAO integrating circuit

By the early 1980s, InSb photodiodes had improved considerably. Their stability was such that DC coupled measurements were possible using an integrating circuit. Barton and Allen [14] described a circuit in which the feedback resistance is replaced with a capacitor, on which the photo-current is integrated. The voltage across the capacitor is measured at the beginning and end of an integration period using a sample and hold circuit, after which the circuit is reset using a brief flash from an LED. From the change in voltage during the integration period, the accumulated photo-current and dark current is measured.

This system circumvents many of the problems associated with the TIA amplifier approach. By varying the integration time, saturation at high photon fluxes is avoided, and the wide range of integration times allows use for both slow chopping ($\sim$ 1 Hz) photometry and rapid observations such as occultations, without loss of high frequency information, as no RC time constant is associated with the circuit. Rapid variations in the DC background cannot be handled by conventional AC lock-in amplifiers, but the Barton and Allen circuitry does not suffer from this problem, allowing them to scan across a whole atmospheric window at 1% spectral resolution in only a few seconds, reducing the effects of poor seeing or thin cloud.

However, the most important aspect of the Barton and Allen circuit is its excellent DC stability, which allows it to be used in raster mapping experiments without frequent sky measurements. Sky measurements are made as required, depending upon the $1/f$ component or sky noise at the wavelength of operation [9].

Based on their circuit, Barton and Allen built an infrared photometer spectrometer (IRPS) for the 3.9 m Anglo-Australian Telescope at Siding Spring, New South Wales. Although the telescope and site are not optimum for true thermal infrared observations, the AAT is a very high quality optical telescope, with an accurate telescope drive that can be used to make precise raster maps. Excellent results have been obtained in the non-thermal infrared between 1 and 3.5 μm with the IRPS.

Many infrared sources have associated low surface brightness extended emission, and DC integrating raster mapping with the IRPS has proven very powerful at making high resolution images of galaxies [6], reflection nebulæ [7], star forming regions [69], and so on. Imaging through broad band filters and narrow band circular variable filters, images of star forming regions in lines of ionised and molecular hydrogen, and nearby continuum have been superimposed using three colour overlay techniques developed by David Malin for optical photography, and spectacular false colour composites have been made [8]. Until the advent of true infrared imaging arrays, the IRPS represented the state-of-the-art in infrared imaging.

2.1.5.2 The Wyoming integrating preamplifier

With a simpler version of the Barton and Allen approach, the astronomers of the Wyoming Infrared Observatory (WIRO) have built a DC integrating system for their 2.3 m telescope located on Jelm Mountain, outside Laramie, Wyoming. The reasons for implementing this system were much the same as for Barton and Allen, again with an emphasis on the DC stability required for raster mapping.

In their circuit, Hackwell, Grasdalen, and Gehrz [56] replaced the Barton and Allen LED reset mechanism with a simple reed relay. Rather than sampling the voltage on the capacitor at the beginning and end of an integration cycle, the Wyoming circuit amplifies and low pass filters the output, and samples it with an analogue-to-digital converter at 1 kHz, synchronised with the raster scanning telescope. Software is used to calculate a least squares fit to the slope of the voltage change on the capacitor, thus deducing the magnitude of the photo-current and dark current. This is in place of the sample and hold circuitry used by Barton and Allen, replacing most of the electronics with software, and equally sensitive results are achieved.

The Wyoming group have developed sophisticated techniques for computer controlled mapping [53], and this system has been used for many imaging experiments both with conventional bolometers [55], in the near infrared, using the DC integrating preamplifier (e.g. [129,34]).

2.2 Infrared array detectors

Despite progress made with DC integrating systems, and accurately controlled telescopes for raster mapping, these systems are still essentially limited by the fact that only one detector is used. There are methods of increasing the efficiency of single element detector systems through the use of multiplex masking [86], but these techniques only gain in the system noise limited regime. With increasingly sensitive detectors, the background limit becomes easier to reach, eliminating the gain of such techniques.

One obvious way to increase observing efficiency and effective sensitivity is to image the scene with many detectors simultaneously. Next, we shall examine the advantages and disadvantages that infrared array detectors have compared to single channel instru-

ments.

2.3 Advantages and disadvantages of array detectors

2.3.1 Advantages

Although the gain associated with using large numbers of detectors simultaneously is evident, the advantages of infrared arrays are many, some more obvious than others. We shall briefly discuss some of them here.

2.3.1.1 Multiplex advantage

For an experiment that requires an image to be made with seeing limited resolution to a given sensitivity limit, an infrared array with $N \times N$ pixels will complete the experiment N^2 times faster than a single element detector, if each pixel of the array is as sensitive as the single element detector. Alternatively, in the background limit, an array of $N \times N$ pixels will reach a sensitivity limit N times fainter than a single element detector, if both use the same total amount of integration time to image a certain fixed area. As arrays are already being manufactured with numbers of pixels in the range 4000–16000, the multiplex advantage is clearly large.

2.3.1.2 Spatial resolution

Whilst it is possible for single element detectors to achieve very high (seeing limited) resolution, it is not simple to make an accurate raster map at this resolution, due to long term telescope drifts and seeing variations. Images made with infrared arrays are immune to such first order drift, as all pixels in the image are seen simultaneously.

2.3.1.3 Geometric stability

An infrared array has a defined physical distance between each of its pixels that can be relied upon to remain fixed during the taking of an image. Thus the relative positional accuracy of sources within one frame can be extremely well defined.

2.3.1.4 On-chip integration

One of the most important qualities associated with infrared arrays is their DC stability, and their ability to integrate flux from the scene until enough background flux has been collected to overcome the system (read) noise. In this way, experiments at almost any spatial or spectral resolution can be made background limited with a long enough on-chip integration time, provided $1/f$ components in the background do not dominate.

2.3.1.5 Simultaneity

Even if drifts in the sky background imply $1/f$ noise, the effect is considerably reduced with infrared arrays, as every pixel is imaging the scene simultaneously. Thus, each detector in the array will see the same total integrated sky flux level at the end of an integration period, as long as the detectors are sufficiently linear, and spatial variations in the sky on scales smaller than the array size are smoothed during the same period. The same advantage applies to working in bad weather, where changes in extinction will be roughly equal for all pixels in the array.

2.3.1.6 Optimal background rejection

Single channel photometers are generally used with apertures large enough to allow for telescope drift and seeing variations. This often results in excess background flux being measured also, reducing the effective signal to noise for point sources. Infrared arrays with small pixels limit the background seen per pixel, and optimum sky subtraction can be done in software after the image is taken, with point source profile fitting or aperture photometry parameters chosen to suit the data *a posteriori* rather than *a priori*. In this way, the maximum signal to noise for point sources is obtained, providing accurate flat fielding can be performed.

2.3.1.7 Off-line optimisation

In the background limited case, small pixels may be binned in software to increase the signal to noise to low surface brightness sources, achieving exactly the same signal to

noise as would be obtained by a single channel photometer with a large beam, also operating in the background limit. As many sources are likely to be structured at some level however, software may be used to optimise spatial resolution in regions of higher intensity, and sensitivity in regions of lower surface brightness, with variable binning. This option is not available with single channel devices. Again, this is a form of *a posteriori* optimisation that is only available with arrays.

2.3.1.8 Large area coverage

The large number of pixels not only implies that images of a fixed area may be made much more quickly than with a single element detector, but also that very large scale survey experiments are possible. That is, we can either consider a 64×64 pixel array to be 4000 times quicker than a single element detector, or 4000 times larger. In addition to surveys, a large simultaneous spatial coverage means that only an approximate location for a source need be known, in the case of IRAS error boxes for example, and it is likely that the source will fall somewhere on the array if the telescope is pointed at the centre of the error box.

2.3.2 Disadvantages

There are some disadvantages to infrared arrays, but most are minor, and may be overcome quite easily.

2.3.2.1 Expensive

Infrared arrays are still relatively expensive, and this prevents them being generally available to the user community. However, many of the major observatories in the world are committed to building infrared camera systems, and the detector array represents a minor fraction of the total cost of building a common-user infrared array camera around one.

2.3.2.2 Software intensive

Relative to conventional infrared systems, array cameras are software intensive, both in terms of the instrument control, and data reduction and analysis. However, this is a minor problem, as optical astronomers are well accustomed to controlling and reducing data from CCD arrays with up to 1000 times as many pixels as the current generation of infrared arrays. The problem does not really lie with the infrared arrays, but more with the infrared astronomers, who are not used to handling such large amounts of data.

2.3.2.3 Electronic intensive

The statement applied to software is true also for electronics. However, the complexity of running infrared arrays may be somewhat greater, due to the hybrid nature of the arrays. Also, at high thermal backgrounds, very fast electronics will be required in order to read out the array before saturation.

2.3.2.4 Read noise limited at low backgrounds

In the case of an extremely low sky background, the small pixels of an array may remain read noise limited at the end of an integration. In this case, pixels cannot be binned up to achieve the same sensitivity as would have been achieved using a single element detector using a large enough beam that it was background limited. For current array detectors, with read noises on the order of 400 e^- RMS, over 1.6×10^5 background photoelectrons must be collected before the detector is background limited. This may take many minutes even for ground based broad band imaging systems, and the problem becomes much worse for high resolution spectrometers and space based systems, where the backgrounds will be orders of magnitude lower.

Much lower read noise array detectors are under development, and these will help in this area. An alternative technique is to use small arrays of very large pixels, specially designed for detecting low surface brightness extended flux in ultra-low backgrounds.

2.3.2.5 Pixel-to-pixel calibration

When making a raster map with a single element detector, no compensation for detector non-uniformity needs to be made. In contrast, the individual detectors of an array will have slightly different responses, and these variations must be compensated for in order to calibrate the image. This is done by flat fielding, i.e. dividing a source image by an image taken of a uniformly illuminated source. This process is discussed at length in this thesis. Users of optical CCDs are well accustomed to the flat fielding process, and the resulting data are only compromised if the pixel-to-pixel variations cannot be completely removed. Current infrared arrays are relatively non-uniform, but can be precisely flat fielded by using a piece of adjacent sky as the uniformly illuminated source. As we shall see later in this thesis, perfect flat fielding is possible with this technique, leaving the data limited by background shot noise alone, as is the case for single element raster maps.

2.4 Infrared array architectures

Arrays of individual detectors with individual preamplifiers can be made, and some early military applications employed thousands of such individually wired detectors [70]. However, in astronomy, individual detectors have rarely been used in arrays bigger than ten or so (e.g. [150]). Questions of reliability and complexity become prohibitive if very many detectors are used in this way.

With the development of the charge-coupled device (CCD) at the beginning of the 1970s, integrated arrays of silicon detectors sensitive to optical radiation quickly became available. Infrared array detectors based on the same principle soon followed [70], by substituting narrower band gap semiconductor materials such as indium antimonide for silicon.

During the 1970s, much development of infrared array technology occurred, largely sponsored by the military. Two basic approaches emerged that are of use in astronomy, namely monolithic arrays, and hybrids. A third technique, using Schottky barrier platinum silicide arrays, has not been of great importance to astronomy due to a very low quantum efficiency, but does exhibit several interesting features which we shall discuss

below.

2.4.1 Monolithic arrays

Similar to an optical CCD, which is fabricated completely out of silicon, both the photon detection process and the subsequent multiplexed read out taking place in the same material, monolithic infrared arrays substitute an infrared sensitive semiconductor for silicon, such as indium antimonide or extrinsic silicon [119]. These so-called IRCCDs proved relatively successful, and some were evaluated for astronomical purposes [97]. However, optimising exotic semiconductor materials such as InSb and HgCdTe for the best imaging performance *and* multiplexing read out circuitry proved difficult, and the monolithic detector has now been largely superseded by the hybrid approach.

2.4.2 Hybrid arrays

Hybrid arrays separate the detection of the infrared photons and the read out of the resulting electrical signals. Highly optimised infrared sensitive detector arrays are mated to silicon multiplexers via metal interconnects, known as 'bump-bonds'. In this way, the detector array and read out mechanism can be optimised separately, and brought together when fully tested [103]. Figure 2.7 shows a schematic representation of a hybrid array, with the infrared sensitive array illuminated from above, and on converting the photons into charge carriers, the signal is injected into the silicon multiplexer beneath through the bump-bonds, where it is accumulated before being read out.

The actual detector part of the infrared sensitive array (e.g. the p-n junction for photovoltaic arrays) must be on the side facing the silicon multiplexer to allow injection of the photocharge into the output circuit. Thus, the photon illumination takes place at the backside of the array, and through the bulk material. If the material is too thick, the photogenerated charge carriers may recombine before diffusing to the front surface, where they are sensed. Thus, in order to ensure high quantum efficiency, the material must mechanically thinned to the correct depth, or made to that depth by epitaxial growth.

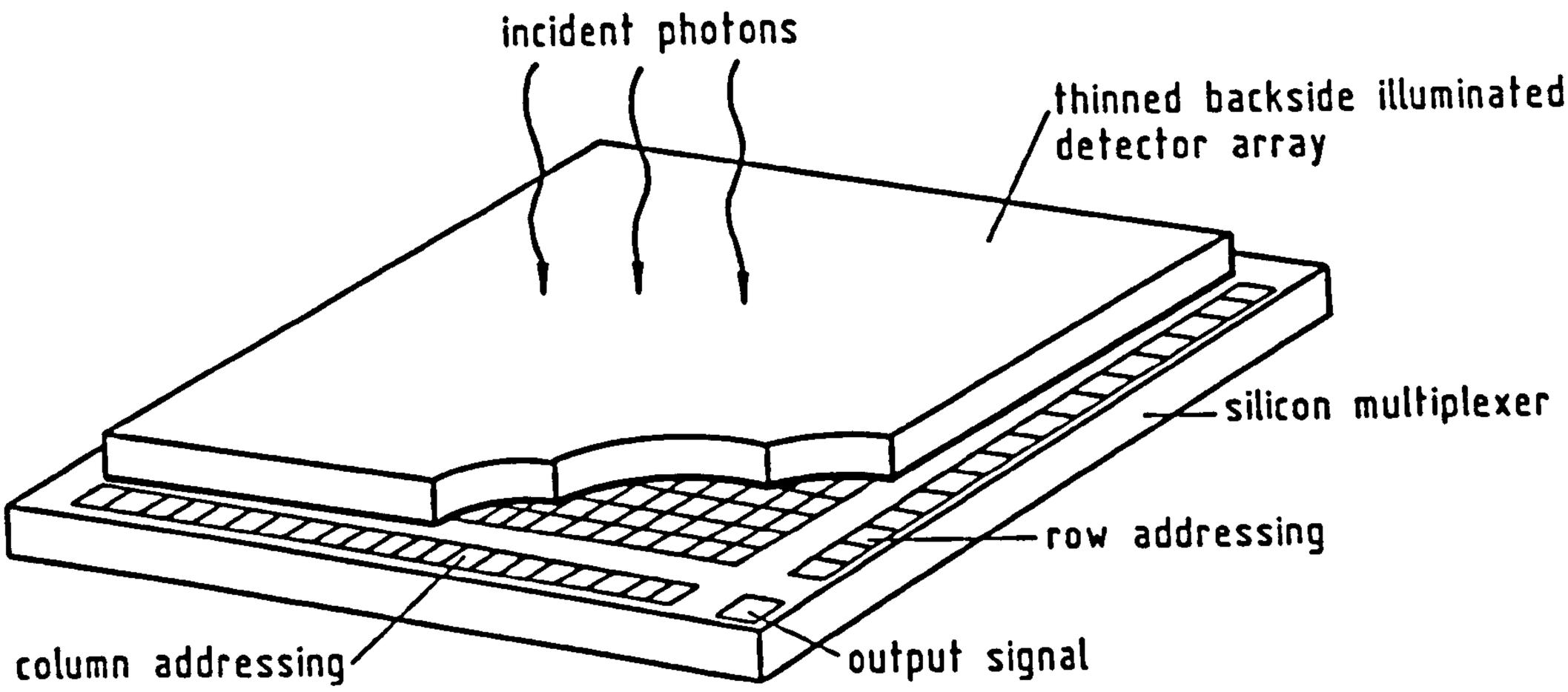


Figure 2.7: Schematic representation of a hybrid array

2.4.2.1 Read-out schemes for hybrids

A range of read out schemes for hybrid arrays is available, including CCDs, charge injection devices (CIDs), switched FET (SWIFET) arrays, and so on. The various approaches are covered at length in the literature, and are well reviewed by Milton [103]. We will only briefly mention the main techniques used.

CCDs have been frequently used, but must be surface channel devices in order to allow the charge generated in the infrared material to be injected into the silicon. These CCDs are often noisier than optical CCDs, which have a buried channel architecture, due to problems associated with surface trapping and so on [38]. However, infrared arrays read out with silicon CCDs have the advantage of having a linear response to photon flux [122], an important factor in their favour for astronomical applications. Large well depths may be obtained by using the CCD as the charge storage site, which is useful under high background conditions [37].

Indium antimonide CID hybrid arrays have been manufactured and array sizes of up to 128 × 128 have been demonstrated [49], and tested under astronomical conditions [40,128]. The arrays tested suffered from various problems, including response

time lag and non-linearity.

Switched FET or (SWIFET), also known as Direct Read Out (DRO) multiplexers are becoming used increasingly [13,66]. These employ an X-Y addressing scheme to read out the detector array, rather than a charge transfer system like that of CCDs. The advantages of this approach include simplicity and compact size, allowing for small area detectors [66]. The circuit is self integrating, and the changing bias during an integration leads to an inherent non-linearity [66]. The SBRC 62 × 58 InSb array is read out via a silicon DRO known as the CRC228, and made by Hughes — this same multiplexer array is used for a range of hybrid type arrays, including the indium antimonide photovoltaic array, and several types of extrinsic silicon photoconductive arrays. We shall be dealing with the SBRC InSb + DRO in detail in chapter 3.

2.4.3 Schottky barrier arrays

Schottky barrier arrays made by platinum silicide (PtSi) and palladium silicide (Pd_2Si). They have not been considered useful to astronomy due to their very low quantum efficiency [80], but have two advantages over other current infrared array technologies. We shall briefly describe their operation [27].

Schottky barrier arrays are made by depositing a thin layer of platinum or palladium onto silicon, forming a silicide layer. The silicide layer is deposited in such a way as to form the electrodes of a CCD. Infrared radiation with wavelengths longer than 1.1 μm enters the backside of the detector and passes through the silicon substrate. The radiation is absorbed in the silicide layer, and charge carriers may be given sufficient energy to cross the metal-silicon Schottky barrier. These so-called 'hot holes' are injected into the silicon substrate as photocurrent. As the holes accumulate in the silicon, a net negative charge accumulates on the silicide, which may be read out via the CCD electrodes.

Thus in this way, the image integrates directly on the output silicide electrodes [27]. The major disadvantage of the technique is that the quantum efficiency is very low, in the range 1–8% [80]. However, two advantages do result. The array is very uniform, as the conversion from infrared photons to charge is almost independent of the doping of the substrate [27]. For real time applications, where no pixel-to-pixel gain or offset

corrections can be made, the Schottky barrier array is highly competitive. Additionally, as the technology is based on that of optical CCDs, large pixel formats have been demonstrated, up to 256 × 256 [78].

2.5 Important detector parameters

In this section, we shall examine the most important parameters that must be considered when an integrated infrared array is considered for use in astronomy. This type of discussion has been covered in the literature, (e.g. [50]). and we shall only briefly summarise the most important points here. We shall refer to the values of the parameters that apply to the SBRC array throughout, in order to assess whether it meets the necessary criteria.

2.5.1 Number of detectors

Since we have decided that arrays of detectors are a 'good thing', we should naturally wish to have as many detectors as possible in our array. Optical CCDs have typical pixel formats in the range 320 × 512 to 800 × 800, with arrays of 2048 × 2048 pixels likely to be available soon [71]. Infrared arrays are some way behind in this regard, with 64 × 64 arrays representing the norm of the current generation. Arrays of 128 × 128 and 256 × 256 pixels are under development by most manufacturers, and are likely to become available soon [13]. Already, the 64 × 64 pixel arrays represent a multiplex advantage of 4000 over the single element detectors still in general use, and there appear to be no fundamental reasons why infrared arrays should not catch up with optical arrays in terms of numbers of pixels.

2.5.2 Read noise

In order that the detector can become limited by the shot noise on the background, the well depth must be large enough that enough background photo-electrons can be collected to overcome the RMS read noise, R_N. This can be trivially stated as :

$$\text{Well depth} \gg R_N{}^2$$

Thus, for a well depth of $10^6\,e^-$, we require a read noise of much less than $1000\,e^-$ RMS. The SBRC array has a well depth close to $10^6\,e^-$ and a read noise noise of $\sim 400\,e^-$ RMS, and thus is background limited once the detector has been filled to about 20%.

2.5.3 Dark current

In order that we may be limited by the shot noise on the sky background and not that on the dark current, we simply require that the dark current be at least two or three times lower than the background photocurrent.

This is easily achieved for ground based instruments, imaging through broad band filters. We shall see later in this thesis that the lowest background photocurrent anticipated for broad band imaging with IRCAM on the UKIRT is $\sim 1500\,e^-$/sec/pixel, and the dark current seen in the InSb arrays is as low as $100\,e^-$/sec/pixel at 35 K. However, in higher spectral resolution imaging, spectroscopy, and space based applications, the background photocurrents may be much lower (10–20 e^-/sec/pixel), and dark currents must be further reduced [67]. Short wavelength cut-off HgCdTe arrays have demonstrated dark currents as low as $2\,e^-$/sec/pixel at 65 K, and for very low background instruments, the detector material cut-off should be carefully chosen to match the longest wavelength of interest, in order not to integrate excess dark current arising due to the unnecessarily small band gap.

2.5.4 Quantum efficiency

Infrared arrays have been developed for a number of non-astronomical purposes, many of which are not fundamentally limited by the brightness of the source of interest, and thus the number of source photons detected per second is not the major concern. In many cases, real time detector uniformity is of greater importance [27]. However, in astronomy, the fundamental sensitivity limit is imposed by the background shot noise, as we are prepared to go to great lengths to correct for pixel-to-pixel non-uniformity. In the background limited case, the signal-to-noise is proportional to the square root of the quantum efficiency, and thus the highest quantum efficiency possible is desired.

High quantum efficiency is available with most near infrared detector materials, at

between 0.3 and 0.7 [50]. Suitable application of an anti-reflection coating to the SBRC InSb array prevents excessive reflection due to the large change in refractive index at the air-detector interface, increasing the peak quantum efficiency to as high as 0.9 (Orias, private communication).

2.5.5 Well depth

The well depth is not of critical importance itself, provided other parameters can be matched correctly. As we have already seen, we require a well depth much largerthan the square of the read noise. The other constraint is that the well does not fill up so fast on thermal background that it cannot be read out before saturating.

The SBRC array has a well depth of $\sim 10^6$ e^-/pixel. At wavelengths beyond 3.6 μm, and at pixel scales of 0.6 $\widehat{''}$, we shall see later that this well depth is not enough to allow a read out before saturation at the current maximum read rate used by IRCAM.

In general, the well depth is linearly proportional to the capacitance of the detector, as is the read noise. Detectors used in future high resolution spectroscopy and space based instruments will have the pixel capacitance reduced in order to reduce the read noise. As the backgrounds are low in these situations, a large well depth is not important. It is important however that the background limit can be reached reasonably fast, to prevent the need for excessive on-chip integrations.

2.5.6 Linearity

For accurate astronomical photometry, we require that the detector be linear in its response to photon flux, or at least calibratable, in the case of non-linear detectors.

Hybrid arrays with CCDs are in principle very linear [122], although this has not always proven the case [38]. The DRO approach is inherently non-linear with respect to photon flux [66], but as we shall see later in this thesis, it is largely calibratable, and should not present a fundamental limitation to the application of DRO arrays to accurate photometric work.

2.5.7 Uniformity

Uniformity of response is important in cases where raw images are to be used. However, as astronomers are willing to process their data off-line, including flat fielding, the importance of inherent non-uniformity is reduced. Nevertheless, it is likely that the fractional non-uniformity left after flat fielding will be proportional in some way to the inherent non-uniformity of the device, and thus the higher the initial uniformity the better.

As we have mentioned, platinum silicide Schottky barrier devices show the lowest inherent non-uniformity, on the order of 0.3% [80]. This is nearly an order of magnitude better than achieved by other infrared array technologies. HgCdTe arrays coupled to CCD multiplexers exhibit non-uniformities on the order of 5% [122]. In the case of hybrid arrays using DRO multiplexers, variations in the quantum efficiency from pixel-to-pixel, and variations in the gains of the individual MOSFETs are responsible for the raw non-uniformity. We shall discuss residual non-uniformity of the SBRC array in detail later in this thesis, and how it is minimised by suitable flat fielding techniques.

2.5.8 Dynamic range

As the range in brightness of astronomical sources is very large, we would like the detector to have a large dynamic range. This parameter is basically specified in terms of the read noise and well depth, the maximum dynamic range for one individual image being given by the well depth divided by the read noise. The dynamic range increases as the square root of the number of images added.

The SBRC array has a well depth of about $10^6\,e^-$, and an RMS read noise of about $400\,e^-$. The maximum dynamic range per image is therefore ~ 2500. In the presence of background flux, the dynamic range is reduced.

2.5.9 Fill factor

An important parameter associated with array detectors is the fill factor : that is, the fraction of the area occupied by each detector that is actually sensitive to illumination.

This fraction should be maximised for the following reasons :

- to increase the effective throughput of the system, by detecting the largest possible fraction of the photons incident on the array
- to minimise the effects on photometric accuracy that occur when point sources are not fully sampled [84]
- to minimise the effects of aliasing that are inherent in an array imaging system [24]

The first of these is obvious; the larger the fraction of the total area that is sensitive to illumination, the higher the throughput, and the greater the total system sensitivity.

If a particular detector design has a low fill factor, other techniques can be employed to minimise the photometric errors at least, especially at low spatial resolution. Having a large image blur relative to the pixel size, photometric errors caused by the positions of stars on the array may be reduced. By defocusing some part of the camera optics, the incoming point source profile may be suitably blurred. However, this technique will probably introduce coma and other field aberrations when a fast lens is defocused, and would produce unsatisfactory images. A second option is that known as micro-scanning. Here, the field of interest is moved relative to the detector array several times in right ascension and declination, by fractions of a pixel each time, effectively blurring the image in a controlled fashion. Micro-scanning may be performed by moving the whole telescope [84], or probably more accurately by careful control of a chopping secondary mirror. The scanning can either take place continuously during an on-chip integration, resulting in a blurred image after read out, or the scanning can take place between on-chip integrations, with a stack of slightly offset images added together later in software. Micro-scanning is better than optical defocusing, as it also helps overcome aliasing, whilst retaining the higher resolution [24]. We shall discuss the problems associated with low fill factor, and the micro-scanning technique that can be used to overcome them further in chapter 4.

Fortunately, a high fill factor is often inherent in a hybrid design infrared array, where the potentially opaque diode and gate structures are on the underside of the chip, and the surface facing the incoming photons is free of obscuration. However, care must be taken that an extremely high fill factor does not lead to excessive pixel-to-pixel

crosstalk, and some small fraction may have to be sacrificed to a perimeter gate, which controls crosstalk. Even with such a gate, the SBRC array has an effective area fill factor in excess of 90% [116].

2.6 The SBRC 62 × 58 InSb + DRO array

Infrared array technologies have matured considerably over the past few years, and the current state of the art is reviewed in a recent edition of Optical Engineering entitled 'Infrared Focal Plane Arrays' [2]. A wide range of these technologies are at last becoming available to the astronomical community. An excellent overview of the detector arrays and the instrumentation based around them is to be found in the proceedings of a recent workshop held in March 1987, in Hilo, Hawaii entitled 'Infrared Astronomy with Arrays' [165].

In the 1–5 μm region, the best array technology currently commercially available is the SBRC 62 × 58 InSb + DRO array. Manufactured by the Santa Barbara Research Center, this array is a hybrid type, with a backside illuminated InSb detector array bump-bonded to an X-Y addressed switched FET output multiplexer. This array is currently in use at UKIRT, NOAO, CTIO, and other sites, and is being considered for a range of new instrumentation, including spectrometers and space based cameras [98,41]. We shall be discussing this array in detail in the following chapter.

2.7 Drives towards new technologies

The development of new infrared array technologies is very rapid. Whilst much of the impetus is still military, an increasing amount of development work for astronomy is being undertaken. We shall briefly discuss these latter drives.

The current astronomical drives are towards highly sensitive array detectors for use in space based instrumentation. Two major cryogenically cooled infrared space telescopes are being developed — NASA plan to fly the Space Infrared Telescope Facility (SIRTF) towards the middle of the 1990s [155], whilst the European Space Agency is already building ISO, the Infrared Space Observatory, due for launch in 1992/3 [76].

Both projects rely heavily on the development of detector arrays for use in the ultra-low background environment of space, and are pushing the limits of the technology far beyond what is required for ground-based use [156,127]. In addition, second generation instruments are already being developed for the Hubble Space Telescope, with an emphasis on short wavelength detector arrays, as the HST is not a cryogenically cooled telescope. Finally, infrared arrays will be required for the proposed Large Deployable Reflector (LDR) space telescope [156].

In all cases, the technology developed will undoubtedly be applied to ground-based astronomy, both during the test phases, and beyond. Also to be considered from the ground-based perspective is the next generation of large telescopes, some already being built, and many in the planning stages. Virtually all of these telescopes are being designed with both optical and infrared instrumentation in mind, and infrared cameras feature in most proposed instrumentation packages [47].

Chapter 3

The SBRC array model

3.1 Introduction

In this chapter we shall discuss some of the semiconductor physics that describes the operation of the SBRC array. The SBRC array is a hybrid type — it comprises a 62×58 array of photo-voltaic diodes implanted into bulk indium antimonide, and a silicon multiplexer chip that employs the direct read-out (DRO) technique (manufactured by Hughes Carlsbad and known as the CRC228). At SBRC, the InSb array is turned over, and bump-bonded to the DRO via indium interconnects. The InSb is illuminated from the back, through the bulk material, and thus must be thin enough to allow charge carriers generated in the bulk to diffuse to the p-n junction, where they are detected. The thinning is done mechanically, after the InSb array has been bonded to the DRO multiplexer. An epoxy resin is used to fill the space between the two arrays, and to relieve the stress caused by the thinning process. Figure 3.1 shows a schematic representation of the physical layout of the SBRC array. The architecture of the SBRC array has consequences for its use in infrared astronomy, most importantly with regards the linearity of the device. The SBRC array has an inherently non-linear response to varying levels of photon illumination, and independently, the dark current generation rate is also non-linear. Eliminating the combination of these two effects is of prime importance to the astronomical user, who is interested in the accurate determination of the relative brightnesses of sources across a large dynamic range.

In order to assess the varying degree of non-linearity, it is useful to develop a software

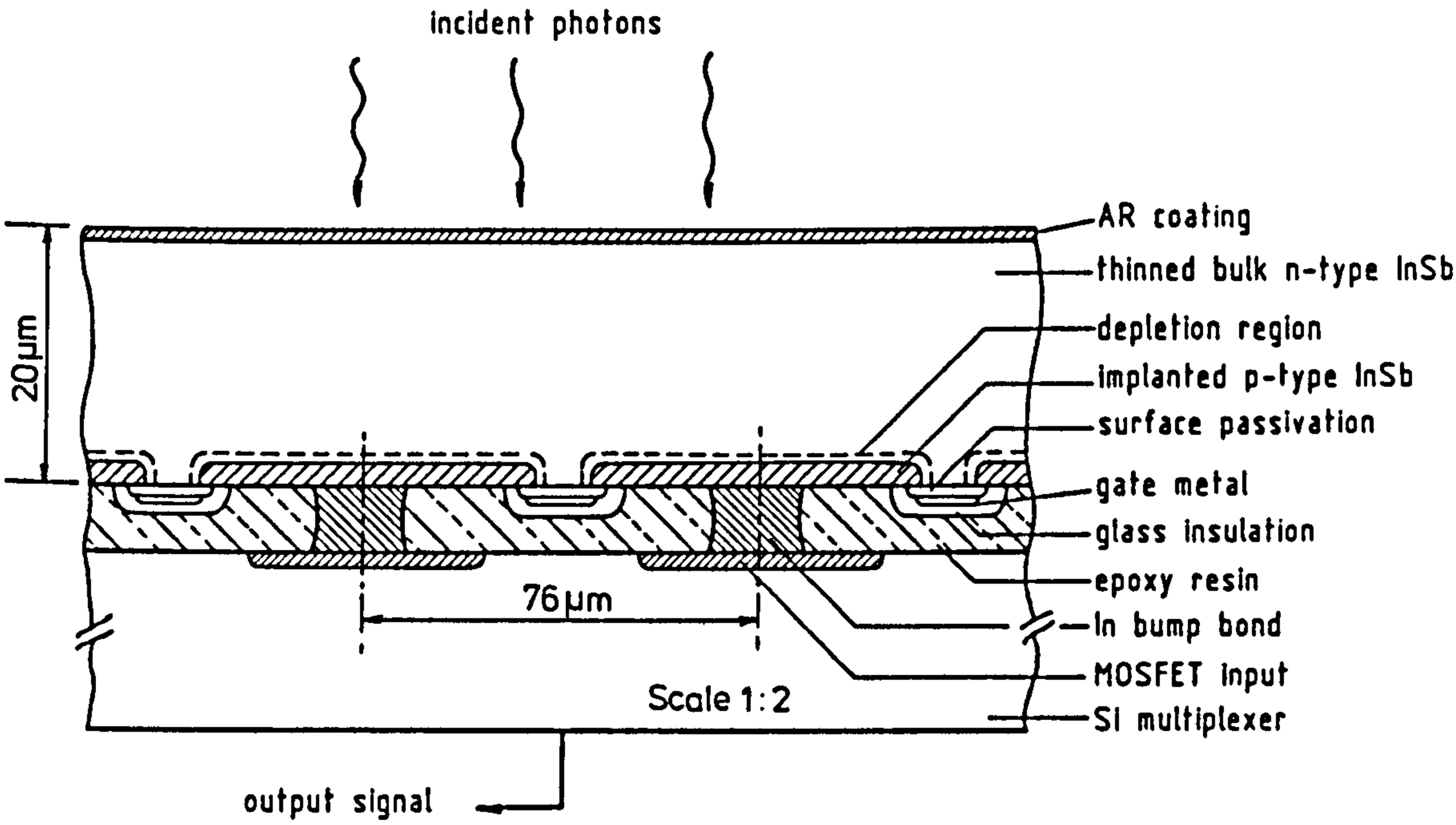


Figure 3.1: Physical layout of the SBRC array

model of the array circuitry, allowing various input parameters to be presented to the model, each set identifying a likely observing regime. Secondly, the development of such a model will aid the search for a suitable linearisation algorithm. Thirdly, such a model can be used in the overall simulation system SIRCAM, as described in chapter 4, as part of the overall goal of deriving predicted system sensitivities.

3.2 Circuit outline

The designers of 2-dimensional arrays of detectors are faced with the same basic problem as applies to all detectors — how to turn the photo-generated current into a measurable output. As we have seen previously, progress in single element detector technology led to the development of DC stable systems, and the use of integrating pre-amplifiers in which current produced in a PV InSb diode is integrated as a voltage ramp on a capacitor. The change in the charge stored between the start and finish of a defined integration period is a measure of the current flowing in the diode, and thence the level of illuminating photon flux during that same integration period.

Given the multiplexed nature of 2-dimensional arrays, it is easily seen that observing

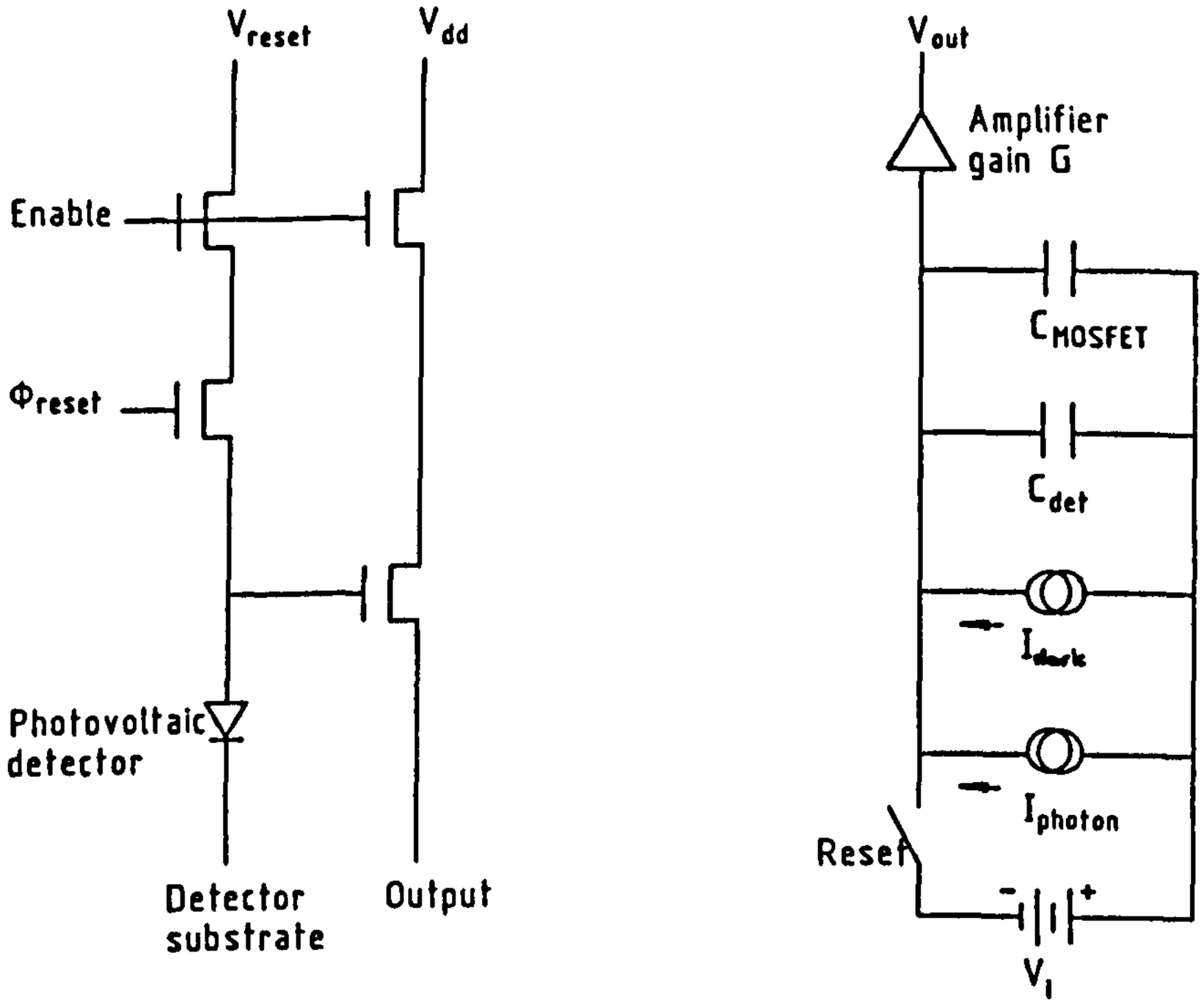


Figure 3.2: Unit cell diagrams for the SBRC array

efficiency is greatly enhanced if the photo-current can be similarly integrated, such that all pixels are simultaneously making measurements during the cycle time necessary to read out the whole array sequentially. The complexity of such integrating circuitry is limited however by the need to provide such a circuit for each diode within the centre to centre spacing of the detectors. Hence the circuitry tends to be simple.

One such simple circuit is known as the Source Follower per Detector (SFD), also known as the Direct Readout (DRO) circuit. This circuit is easy to operate and will fit into a suitably small area ($\leq 40\,\mu$m [66]). The particular DRO circuit circuit used to readout the SBRC array is the CRC228, developed by Hughes Carlsbad Research Center, a sister division of SBRC. Figure 3.2 shows the unit cell circuit for InSb diode array and CRC228, and the equivalent unit cell diagram we shall use whilst developing our detector model.

Examining these figures, the theory of operation may be seen. At the beginning of an integration period, the reset switch is closed, back biasing the detector to an initial voltage V_i, and charging up the effective capacitances C_{det} and C_{FET}. The reset switch is then opened. The currents generated in the diode are I_{photon}, due to incoming photons, and I_{dark}, the dark current. These act to discharge the capacitance. After a certain

integration time the output voltage, V_{final}, is sampled, indicating the voltage remaining on the circuit. The detectors may be reset at this point to start a new integration. There is a non-destructive read out option, where the voltage is sampled without resetting the detector. Taking the difference between the voltages sampled at the beginning and end of the integration period should, in principle, allow the magnitude of $I_{photon} + I_{dark}$ to be deduced. It should be noted that two output lines are used in the CRC228, such that pairs of columns are read out simultaneously.

However, as we shall see, both I_{dark} and C_{det} are functions of the bias voltage V_i, and by definition in this circuit, the bias voltage changes during an integration. Thus the discharge of the total node capacitance is not trivially described.

We now examine the equation of discharge for the unit cell. We are concerned with the voltage sampled at the MOSFET, and thus with the rate of change of this voltage, dV/dt. This work follows from previous work by Vampola [149] for other detector systems, and additional work by Minshull and Botts [104], all at SBRC.

The charge Q on the circuit is given by the product of the capacitance and voltage :

$$Q = C \times V \tag{3.1}$$

By differentiation we obtain :

$$\begin{aligned} dQ &= CdV + VdC \\ &= CdV + V\frac{\partial C}{\partial V}dV \\ &= \left(C + V\frac{\partial C}{\partial V}\right)dV \end{aligned} \tag{3.2}$$

The change in the charge, dQ, is due to the flow of a current I_{det}, for a time dt. Thus :

$$dQ = I_{det}dt \tag{3.3}$$

Combining these equations, we obtain

$$-\frac{dV}{dt} = \frac{I_{det}}{\left(C + V\frac{\partial C}{\partial V}\right)} \tag{3.4}$$

This is the basic 'equation of motion' for the unit cell discharge. We next investigate the nature of the components of this equation. We have adopted the convention that a reverse bias has negative sign. As we shall see below, in reverse bias, the detector current has negative sign. Thus, in order to make the reverse bias discharge towards zero, we

require the signs as shown. Throughout the following analysis, we shall maintain the convention that a reverse bias voltage has a negative sign.

3.3 Detector current

The detector current, I_{det}, has two components, namely I_{dark}, the dark current, and I_{photon}, the photo-generated current. The use of a minus sign rather than a plus sign here is explained later (see page 72).

$$I_{det} = I_{dark} - I_{photon} \tag{3.5}$$

3.3.1 Dark current

The dark current, I_{dark}, is seen in the absence of incoming photon illumination, and is due to thermal generation of charge carriers within the semiconductor. We shall analyse the two main components of the dark current, namely the diffusion current and the generation-recombination (G-R) current. We shall specifically exclude other sources of dark current, including general surface effects, and bulk and surface tunnelling currents. These latter effects are hard to treat analytically, but may in fact become the dominant sources of dark current below 40 K in InSb [142]. Section 3.3.2 briefly examines these other current sources, and derives an order of magnitude estimate for one, namely tunneling current.

3.3.1.1 Diffusion current

Diffusion current arises when carriers are generated or recombine in the bulk semiconductor material, away from the p-n junction. Without the influence of an electric field to move them, the carriers diffuse through the material. In a detector under reverse bias, minority carriers arriving near the depletion region will then be swept across the junction by the electric field present there, generating a reverse current. Under forward bias, majority carriers arriving at the junction will see a reduced potential barrier, and they will be swept across the junction. Arriving at the other side of the junction, the majority carriers now become minority carriers; they combine with majority carriers

supplied at the front contact, and result in a forward current.

The diffusion current may be written as [149] :

$$I_{diff} = I_{0_{diff}}(e^{V/V_t} - 1) \tag{3.6}$$

where V is the detector bias, V_t is the 'thermal voltage' (kT/q), and $I_{0_{diff}}$ is given by :

$$I_{0_{diff}} = A_{det} q n_i^2 \sqrt{V_t} \left(\frac{(\mu_n/\tau_n)^{\frac{1}{2}}}{N_A} + \frac{(\mu_p/\tau_p)^{\frac{1}{2}}}{N_D} \right) \tag{3.7}$$

where A_{det} is the detector area

n_i is the intrinsic carrier concentration

k is Boltzmann's constant

T is the detector temperature

q is the electronic charge

$\mu_n = D_n/V_t$, where D_n is the minority carrier diffusivity on the n side of the p-n junction

$\mu_p = D_p/V_t$, where D_p is the minority carrier diffusivity on the p side of the p-n junction

τ_n is the minority carrier lifetime on the n side

τ_p is the minority carrier lifetime on the p side

N_D is the ionized donor concentration on the n side

N_A is the ionized donor concentration on the p side

These equations show the functional dependence of the diffusion current. Of note is that :

- $I_{0_{diff}}$ depends on $\mu_n, \mu_p, \tau_n, \tau_p$—these parameters are process dependent and difficult to control

- $I_{0_{diff}}$ depends on N_D, N_A—these parameters are optimisable during processing, and may be altered in order to reduce the dark current

- $I_{0_{diff}}$ is proportional to n_i^2, the square of the intrinsic carrier concentration—this is an important point which we shall return to later

3.3.1.2 Generation-recombination (G-R) current

Generation-recombination, or G-R current arises in the depletion region of the p-n junction, due to the emission and capture of charge carriers through mid-band traps (see [167, page 98] and [134, pages 35,89]). Both the minority and majority carriers are detected, as they are swept out of the depletion region in opposite directions by the electric field. The general form of the current is given by [149] :

$$I_{G-R} = I_{0_{G-R}}(1 - V/V_{bi})^{\frac{1}{2}}(e^{V/2V_t} - 1) \tag{3.8}$$

with $I_{0_{G-R}}$ given by :

$$I_{0_{G-R}} = \frac{A_{det} q n_i \sqrt{V_{bi}}}{2\tau_0}\left[\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{\frac{1}{2}} \tag{3.9}$$

where V_{bi} is the diode built-in voltage (see below)

τ_0 is the effective depletion region minority carrier lifetime

ϵ_s is the static dielectric constant for the given semiconductor material ($= 17.7\,\epsilon_0$ for InSb)

ϵ_0 is the permittivity of free space ($= 8.854 \times 10^{-12}$ F/m)

Note the similarity between the form of the G-R current and the previous equation for diffusion current. The extra factors of two that appear in the G-R current are due to the fact that both the minority and majority carriers are detected, as opposed to only one or other, depending on the sign of the bias, being sensed in diffusion current. The additional term $(1 - V/V_{bi})^{1/2}$ in the G-R current is due to the fact that it arises in the depletion region. Thus the magnitude of the G-R current will depend on the volume of the depletion region, and thence, for a fixed detector area, the width of the depletion region. This width, W, of an abrupt junction is given by [134, page 529] :

$$W = \left[\frac{2\epsilon_s}{q} V_{bi}(1 - V/V_{bi})\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{\frac{1}{2}} \tag{3.10}$$

As we shall see later, the typical SBRC doping gives an acceptor doping level (N_A) about 10^4 times larger than the donor doping level (N_D), and it is safe to assume the abrupt junction approximation in these circumstances. The width of a junction with these doping levels is around $1\,\mu$m at a reverse bias of -250 mV. This equation gives us the $(1 - V/V_{bi})^{1/2}$ dependence of the G-R current. Note also the critical dissimilarity between the form of the diffusion and G-R currents — the former is proportional to n_i^2, and the latter to n_i only. This will be shown to be of great importance later.

3.3.1.3 Total dark current

Combining the diffusion and G-R currents, we obtain the total dark current, I_{dark} :

$$I_{dark} = I_{0_{diff}}(e^{V/V_t} - 1) + I_{0_{G-R}}(1 - V/V_{bi})^{\frac{1}{2}}(e^{V/2V_t} - 1) \tag{3.11}$$

Next we shall rewrite the forms of $I_{0_{diff}}$ and $I_{0_{G-R}}$ in more accessible terms, in particular R_0, the dynamic detector impedance.

3.3.1.4 Dynamic detector impedance

The dynamic detector impedance at zero bias, or R_0, is obtained by evaluating the derivative of the the detector current I with respect to the voltage V at zero bias thus :

$$R_0{}^{-1} = \left.\frac{dI}{dV}\right|_{V=0} \tag{3.12}$$

Differentiating equation 3.11, with respect to V, and evaluating the result at zero bias, we obtain :

$$\begin{aligned} \left.\frac{dI}{dV}\right|_{V=0} &= \frac{I_{0_{diff}}}{V_t} + \frac{I_{0_{G-R}}}{2V_t} \\ &= \frac{1}{V_t}\left(I_{0_{diff}} + I_{0_{G-R}}/2\right) \end{aligned} \tag{3.13}$$

Thus we obtain R_0, the dynamic detector impedance at zero bias as :

$$R_0 = \frac{2V_t}{(2I_{0_{diff}} + I_{0_{G-R}})} \tag{3.14}$$

Now, jumping ahead slightly, we will take it that R_0 is a highly temperature sensitive parameter. As we shall see, R_0 depends on n_i, the intrinsic charge carrier concentration, which as we shall also see, has a strong temperature dependence. As noted earlier, the two dark current mechanisms have different dependencies on n_i, and thus different dependences on the detector temperature. We shall now introduce the concept of two dynamic detector impedances, one each for diffusion and G-R current. The combined R_0 will be a superposition of the two in parallel. We will see that there are two separate temperature regimes in which diffusion and G-R current dominate, and within these regimes, the combined R_0 will reduce effectively to that of the dominant current. Thus, for the two regimes, we obtain :

$$R_{0_{diff}} = \frac{V_t}{I_{0_{diff}}} \tag{3.15}$$

and

$$R_{0_{G-R}} = \frac{2V_t}{I_{0_{G-R}}} \tag{3.16}$$

Rearranging and substituting into equations 3.6 and 3.8, we have :

$$I_{diff} = \frac{V_t}{R_{0_{diff}}}(e^{V/V_t} - 1) \tag{3.17}$$

$$I_{G-R} = \frac{2V_t}{R_{0_{G-R}}}(1 - V/V_{bi})^{\frac{1}{2}}(e^{V/2V_t} - 1) \tag{3.18}$$

3.3.1.5 Temperature dependence

We have already mentioned that the dark current is dependent on the temperature, and we shall now examine this topic more rigorously.

In equations 3.17 and 3.18, the temperature dependent parameters are V_t, V_{bi}, and most importantly, R_0. We see why when we examine the explicit forms of these parameters. The 'thermal voltage', V_t, is defined as :

$$V_t = kT/q \tag{3.19}$$

The diode built-in potential, V_{bi}, is given by :

$$V_{bi} = V_t \ln\left(\frac{N_A N_D}{{n_i}^2}\right) \tag{3.20}$$

This is the built-in potential for an abrupt junction rather than a continuous junction (see [134, page 75]), applicable to SBRC diodes as mentioned previously. By rearranging equations 3.7, 3.9, 3.15, and 3.16, we obtain the two dynamic detector impedances :

$$R_{0_{diff}} = \frac{{V_t}^{\frac{1}{2}}}{A_{det} q {n_i}^2}\left[\frac{(\mu_n/\tau_n)^{\frac{1}{2}}}{N_A} + \frac{(\mu_p/\tau_p)^{\frac{1}{2}}}{N_D}\right]^{-1} \tag{3.21}$$

$$R_{0_{G-R}} = \frac{4V_t\tau_0}{A_{det} q n_i {V_{bi}}^{\frac{1}{2}}}\left[\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{-\frac{1}{2}} \tag{3.22}$$

Examining these four equations, we see that V_t has a linear dependence on the temperature T; V_{bi} depends on T via V_t and $\ln(1/{n_i}^2)$ — these two dependences roughly cancel (see figure 3.4). As mentioned earlier, $R_{0_{diff}}$ has has a dependence on $1/{n_i}^2$, whereas $R_{0_{G-R}}$ depends on $1/n_i$ only. As we shall now see, it is the intrinsic carrier concentration, n_i, that is the critical parameter in controlling the two R_0 s, and thus the two corresponding dark currents, due to a very strong temperature dependence.

The intrinsic carrier concentration, n_i, is given by Yang [167, pages 12 and 13] as :

$$n_i^2 = 2\left(\frac{2\pi m_n^* kT}{h^2}\right)^{\frac{3}{2}} 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{\frac{3}{2}} e^{(-E_g/kT)} \tag{3.23}$$

where m_n^* is the effective mass of the minority carriers on the n side of the junction
(electrons)

m_p^* is the effective mass of the minority carriers on the p side of the junction
(holes)

E_g is the semiconductor bandgap

h is Planck's constant

The effective mass constants for InSb are [134, pages 847,849] :

$m_n^* = 0.015\ m_e$

$m_p^* = 0.390\ m_e$

where m_e is the mass of an electron, 9.11×10^{-31} kg.

It is worth noting here that the bandgap of semiconductors is, in general, slightly temperature dependent, often represented empirically with an equation of the form :

$$E_g = E_{g0} + \beta T \tag{3.24}$$

For indium antimonide the empirically determined constants E_{g0} and β are 0.24 eV and -2×10^{-4} eV respectively (where 1 eV $\equiv 1.6 \times 10^{-19}$ J) [149]. Using these values, we can derive the intrinsic charge carrier concentration, n_i, for InSb as a function of temperature, T, including the temperature dependence of E_g, thus :

$$n_i = 3.26 \times 10^{14} T^{\frac{3}{2}} e^{(-1400/T)} \quad \left[\text{cm}^{-3}\right] \tag{3.25}$$

Note the use of the unit cm^{-3} for n_i — it is not an SI unit, but is used here, as it is the semiconductor industry standard. Figure 3.3 shows the steep temperature dependence of n_i for InSb across the temperature region in which the SBRC array is typically operated.

Now we can substitute for n_i in equations 3.21 and 3.22, in order to examine the explicit temperature dependence of $R_{0_{diff}}$ and $R_{0_{G-R}}$ for InSb.

$$R_{0_{diff}} = \frac{(k/q^3)^{\frac{1}{2}}}{1.06 \times 10^{29} A_{det} T^{\frac{1}{2}} e^{(-2800/T)}} \left[\frac{(\mu_n/\tau_n)^{\frac{1}{2}}}{N_A} + \frac{(\mu_p/\tau_p)^{\frac{1}{2}}}{N_D}\right]^{-1} \tag{3.26}$$

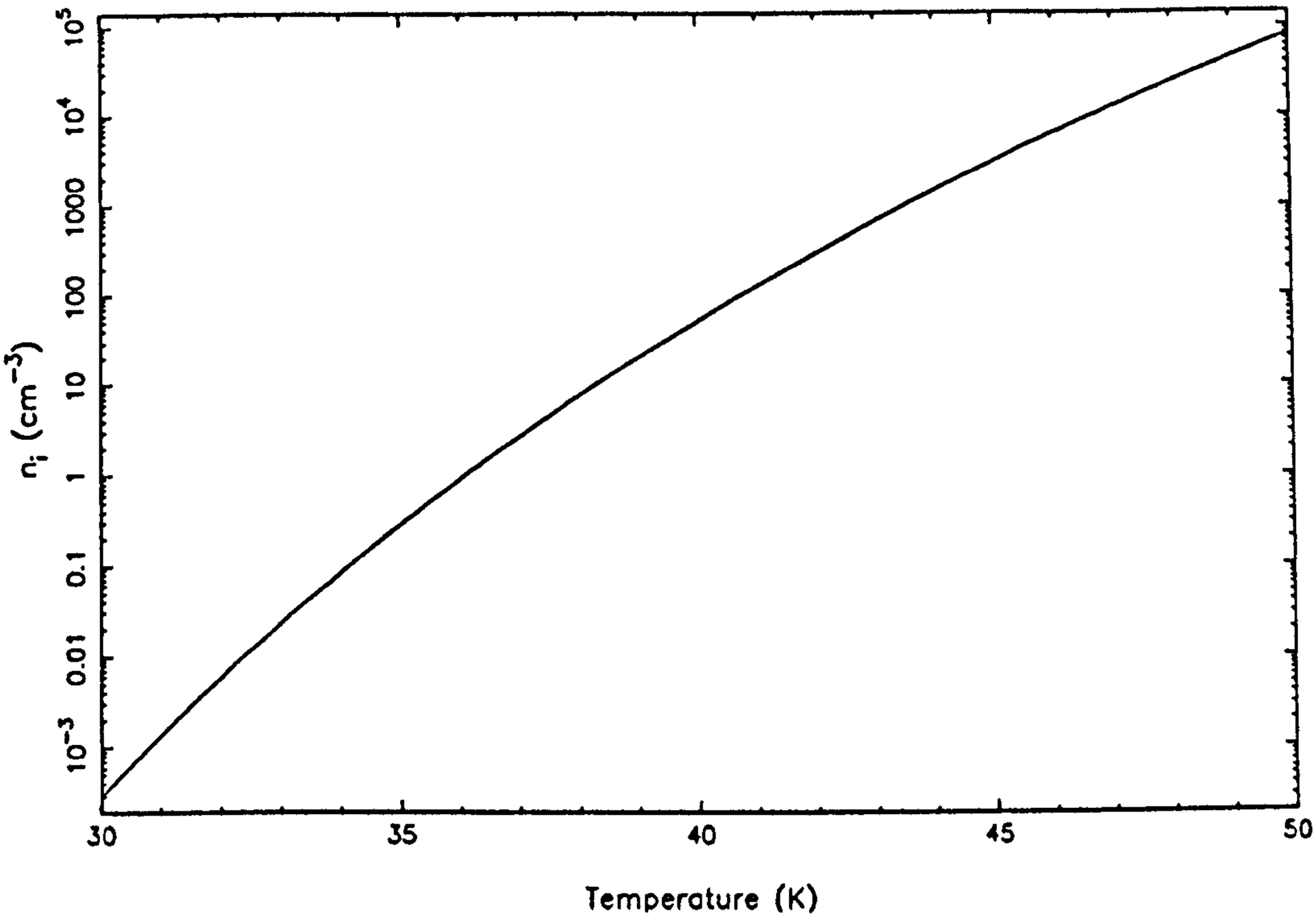


Figure 3.3: Temperature dependence of the InSb intrinsic carrier concentration

$$R_{0_{G-R}} = \frac{4k\tau_0}{3.26 \times 10^{14} A_{det} q^2 V_{bi}^{\frac{1}{2}} T^{\frac{1}{2}} e^{(-1400/T)}} \left[\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{-\frac{1}{2}} \quad (3.27)$$

Although there is a temperature dependence in V_{bi}, it is relatively slight. Figure 3.4 shows the variation of built-in diode voltage V_{bi} with temperature for an InSb p-n junction with typical SBRC doping levels of $N_A = 10^{19}\,\mathrm{cm}^{-3}$ and $N_D = 10^{15}\,\mathrm{cm}^{-3}$. It can be seen that V_{bi} does not change by more than 20% across a wide range of operating temperatures — as the V_t term in equation 3.20 decreases with temperature, the $\ln(N_A N_D / {n_i}^2)$ term increases, with approximate cancellation between the two.

Thus it is the temperature dependencies arising from n_i that are dominant in determining the changes in dark current with temperature. We can collect the essentially constant terms in order to simplify equations 3.26 and 3.27.

$$R_{0_{diff}} = K_{diff} \times T^{-\frac{1}{2}} e^{(2800/T)} \quad (3.28)$$

$$R_{0_{G-R}} = K_{G-R} \times T^{-\frac{1}{2}} e^{(1400/T)} \quad (3.29)$$

The values of K_{diff} and K_{G-R} are approximately constant with temperature and are given by :

$$K_{diff} = \frac{(k/q^3)^{\frac{1}{2}}}{1.06 \times 10^{29} A_{det}} \left[\frac{(\mu_n/\tau_n)^{\frac{1}{2}}}{N_A} + \frac{(\mu_p/\tau_p)^{\frac{1}{2}}}{N_D}\right]^{-1}$$

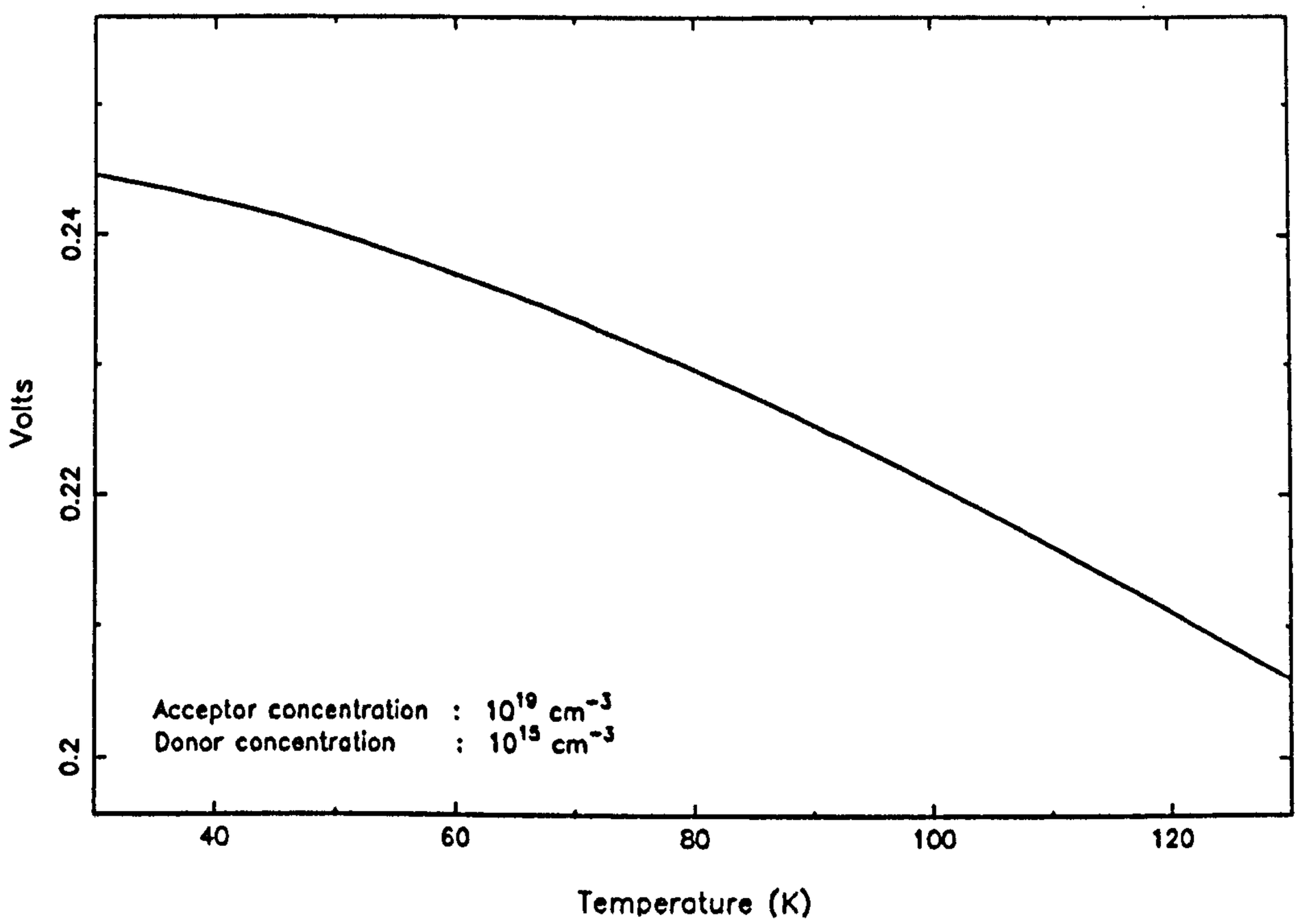


Figure 3.4: Temperature dependence of the built-in diode voltage

$$K_{G-R} = \frac{4k\tau_0}{3.26 \times 10^{14} A_{det} q^2 V_{bi}^{\frac{1}{2}}} \left[\frac{2\epsilon_s}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{-\frac{1}{2}}$$

However, we rarely have sufficiently detailed knowledge of the manufacture and processing of the semiconductor material to be able to calculate the values of K_{diff} and K_{G-R} from the material constants, as given above. Rather, we tend to make measurements of the actual detector R_0 at a few fixed temperatures, and then use the previous equations to calculate the value of R_0 for other temperatures as follows :

$$R_{0_{diff}}(T) = \left(\frac{R_{0_{diff}}(T_0)}{T_0^{-\frac{5}{2}} e^{(2800/T_0)}}\right) T^{-\frac{5}{2}} e^{(2800/T)} \tag{3.30}$$

$$R_{0_{G-R}}(T) = \left(\frac{R_{0_{G-R}}(T_0)}{T_0^{-\frac{1}{2}} e^{(1400/T_0)}}\right) T^{-\frac{1}{2}} e^{(1400/T)} \tag{3.31}$$

where $R_{0_{diff}}(T_0)$ and $R_{0_{G-R}}(T_0)$ are the values of $R_{0_{diff}}$ and $R_{0_{G-R}}$ measured at a fixed temperature T_0. In practice, this entails taking advantage of the fact that diffusion and G-R current dominate in separate temperature regimes as mentioned earlier, and that measuring the total R_0 in either regime will effectively measure just the dominant impedance for that regime.

Figure 3.5 shows the cumulative R_0 for InSb across a range of temperatures, assuming values of $R_{0_{diff}} = 2 \times 10^5\ \Omega$/pixel at 140 K and $R_{0_{G-R}} = 10^{10}\ \Omega$/pixel at 77 K.

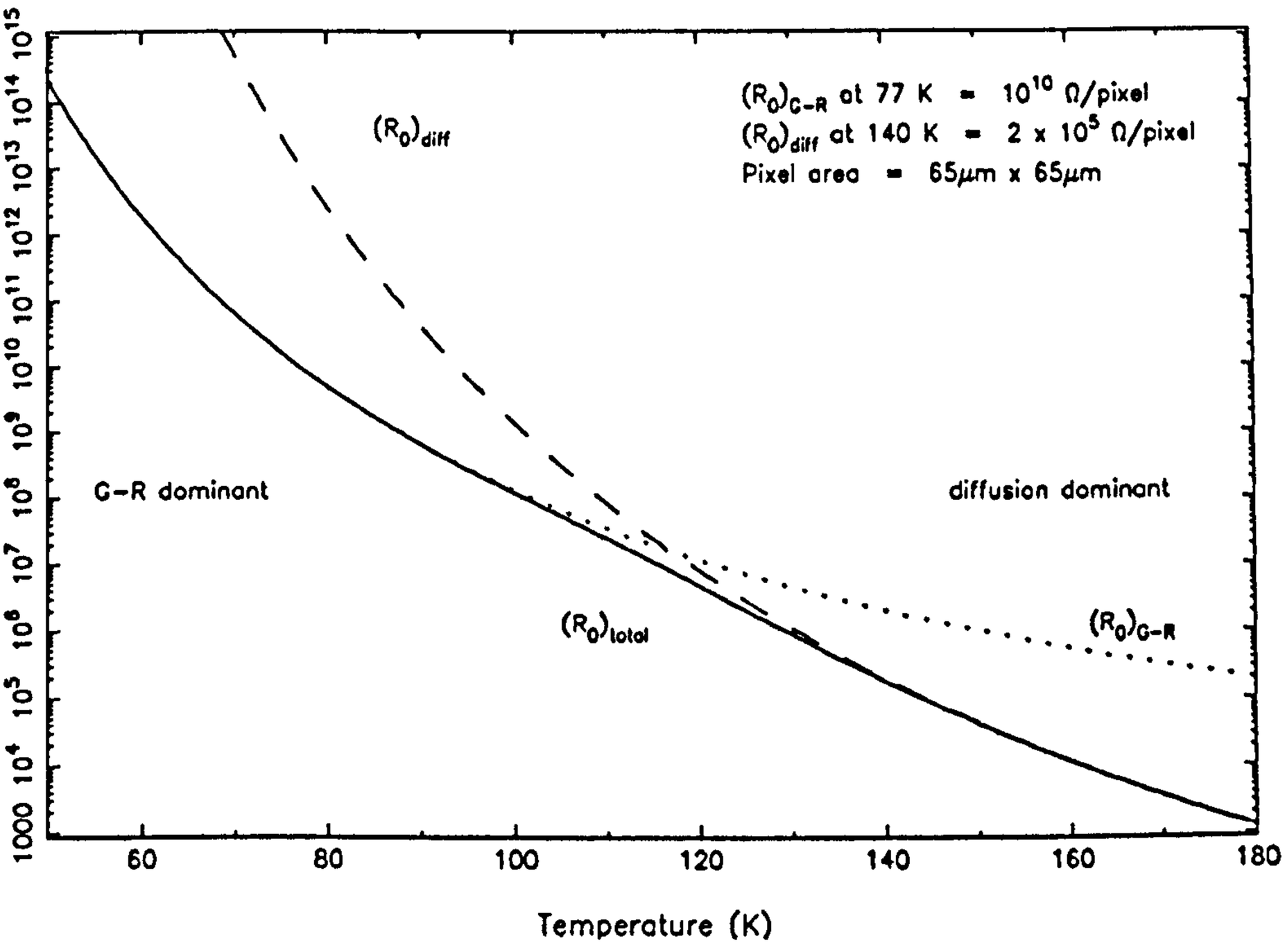


Figure 3.5: Temperature dependence of the dynamic detector impedances

These values are typical of the R_0 s measured for SBRC processed InSb diodes of area $65 \times 65\,\mu$m [149,142]. Note that the two impedances add in parallel :

$$\frac{1}{R_0} = \frac{1}{R_{0_{diff}}} + \frac{1}{R_{0_{G-R}}} \tag{3.32}$$

Figure 3.6 shows the relative contributions to the dark current by the diffusion and G-R currents as a function of temperature for the same detector parameters. We can see that above about 130 K we are dominated by diffusion current. As the detector is cooled, the charge carrier concentration, n_i, drops off steeply, and the diffusion current, depending on n_i^2, decreases more rapidly than the G-R current, which depends on n_i. Between 110 K and 40 K, it is G-R current that dominates. Figure 3.7 shows the predicted number of G-R dark current electrons generated per second as a function of temperature in an SBRC InSb diode, with a typical $R_{0_{G-R}}$ at 77 K of $10^{10}\,\Omega$/pixel, and at a typical mid-well reverse bias of −100 mV. We can see that the model predicts currents on the order of $10^7\,e^-$/second/pixel at 77 K, and for a typical well-depth of $10^6\,e^-$, the array would have to be read-out ten times a second in order to prevent the diodes saturating. Thus, operating the array at liquid nitrogen temperatures (77 K) would be highly unsatisfactory, and lower temperatures must be used. Around 60 K, the dark current is decreasing dramatically, by about an order of magnitude for every

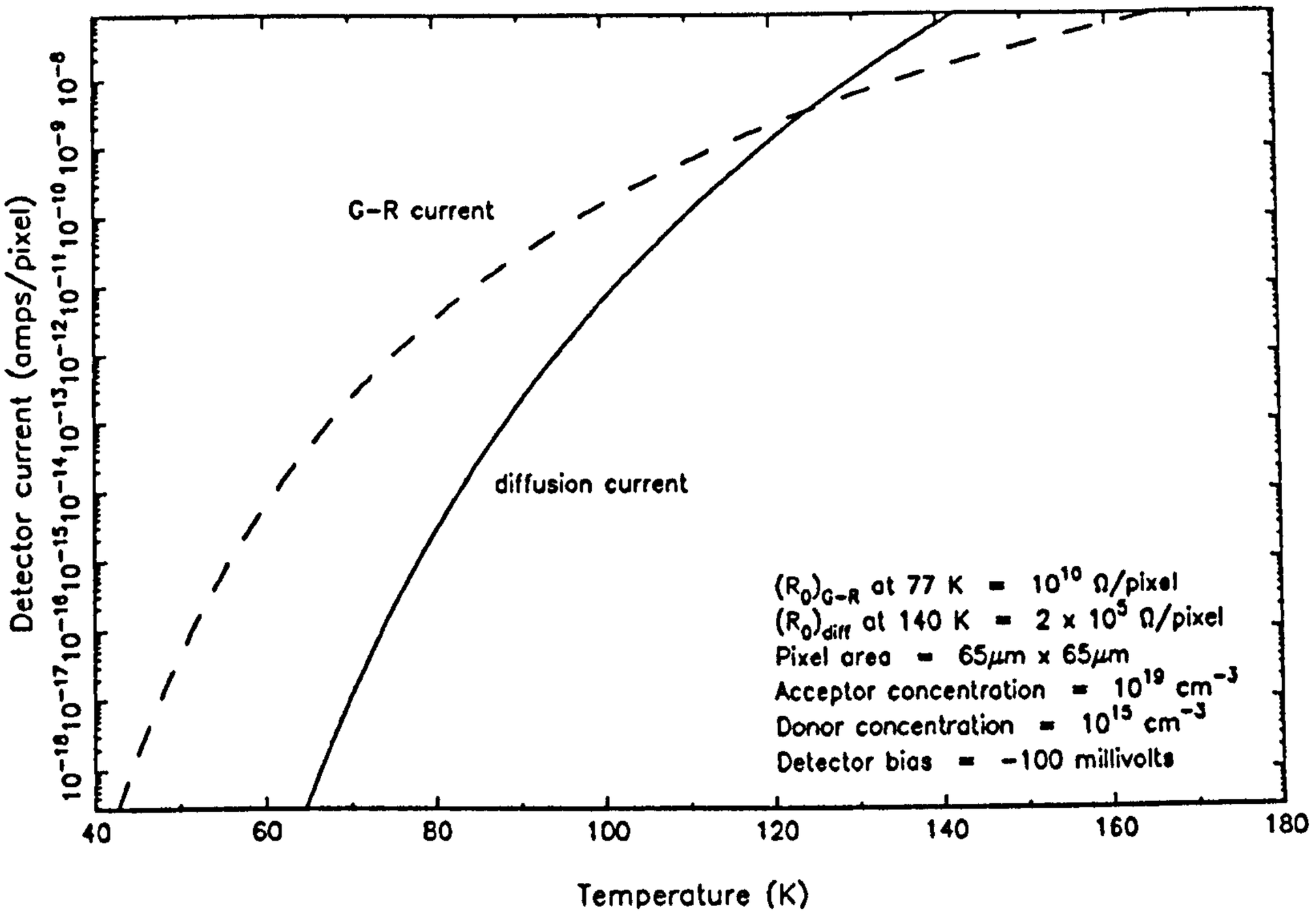


Figure 3.6: Temperature dependence of the dark current

5 K temperature drop.

3.3.2 Low temperature dark current sources

If the dark current were due to generation-recombination and diffusion mechanisms alone, then figure 3.7 shows us that it should be less than 1 e^-/sec/pix at 40 K. However, it is known that significantly larger dark currents are present at this temperature, and well below. At 35 K, the actual dark current is on the order of 100 e^-/s/pix at a reverse bias of 250 mV (McLean, private communication), and SBRC have measured dark currents in the range 25–125 e^-/s/pix at 7 K (Hoffman, private communication). To account for the dark current observed at these low temperatures, we must find mechanisms other than diffusion and generation-recombination.

Such a dark current source should be non-thermal, and roughly independent of detector temperature, given the approximate constancy between 35 and 7 K. Several possible non-thermal sources can be identified in this particular kind of detector array [142] :

- band-to-band tunneling

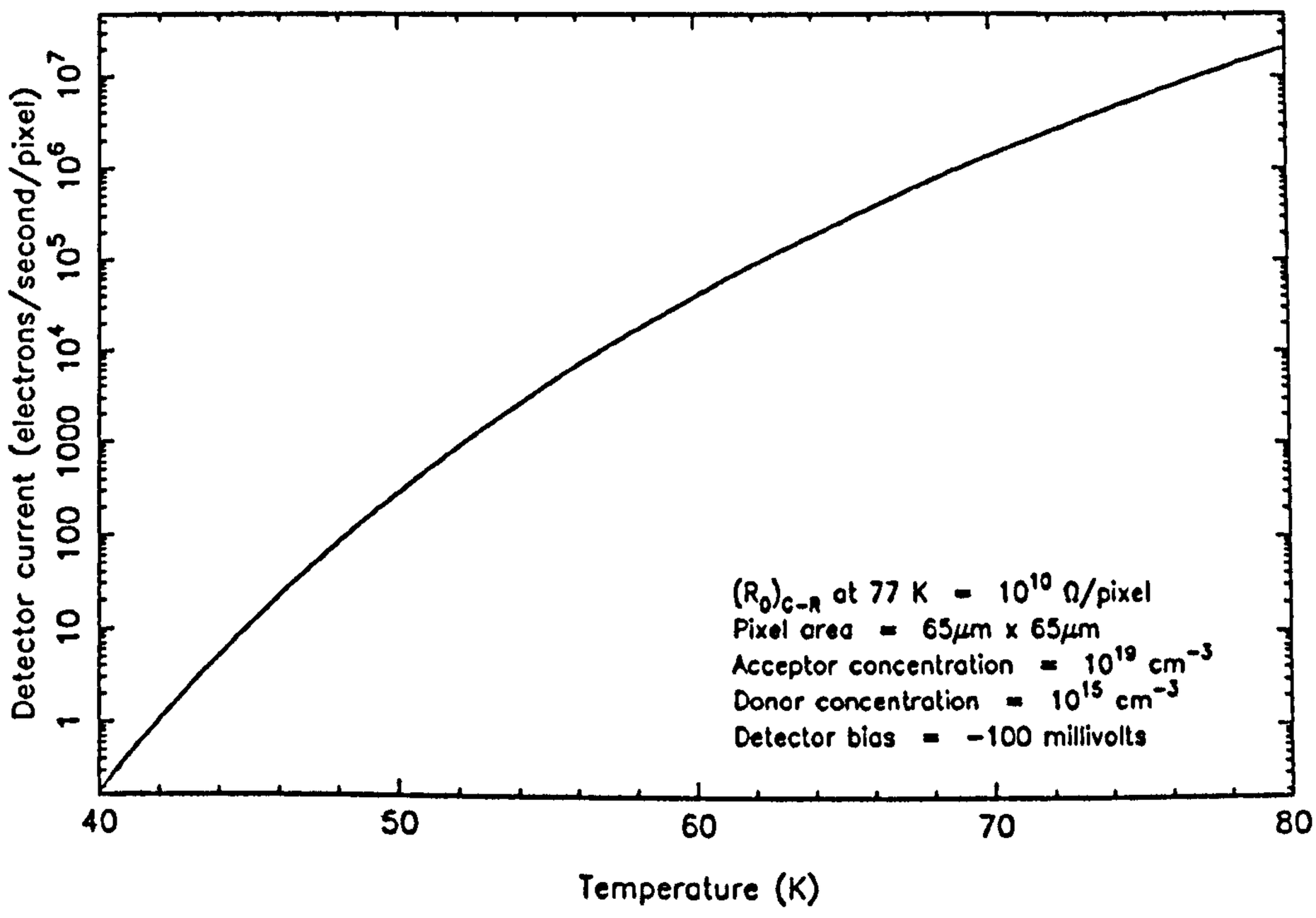


Figure 3.7: Temperature dependence of the G-R current in InSb

- trap-assisted tunneling
- surface region tunneling
- surface leakage
- leakage through the gate oxide

Although the basic physics behind these mechanisms is understood, they cannot easily be characterised, as their magnitudes are critically dependent upon the material parameters and fabrication design of a given detector. However, following the work of Thom and Yang [142], we can at least examine the first of these mechanisms, band-to-band tunneling current, and attempt to discover whether the observed dark current is consistent in magnitude and behaviour with that mechanism.

Band-to-band tunneling is a quantum phenomenon, and detailed discussions are to be found elsewhere [167, p. 100][134, p. 513]. Briefly, tunneling occurs in heavily doped p-n junctions, where the Fermi level lies within the conduction or valence band, and when there are electrons in the n-type conduction band with the same energy as empty states in the p-type valence band, as shown in figure 3.8 for forward bias. If the

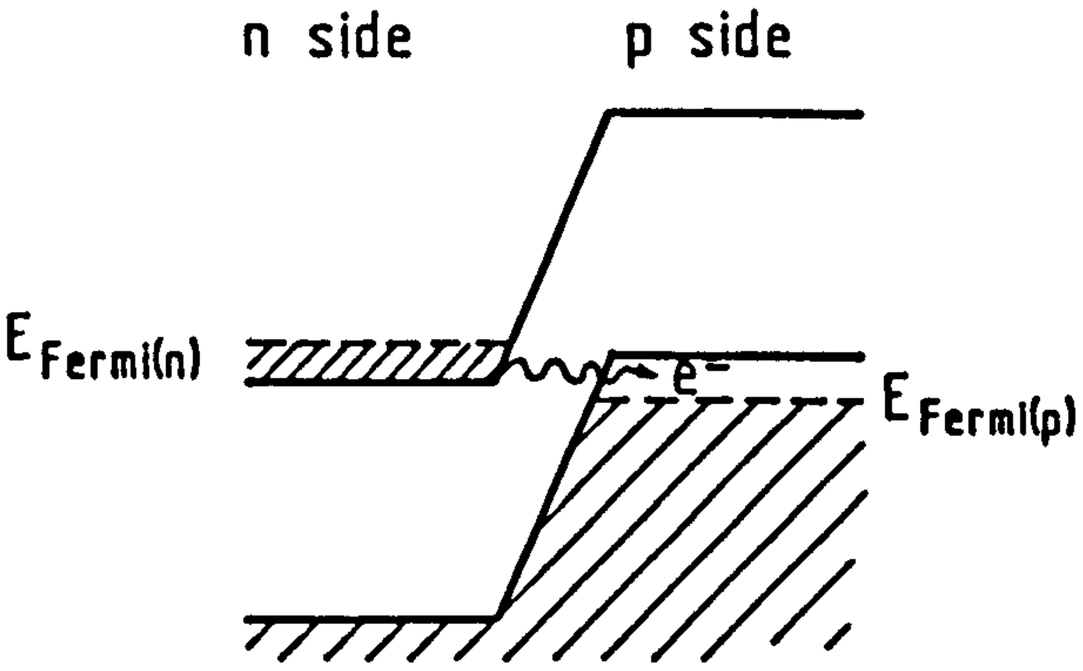


Figure 3.8: Forward tunnelling

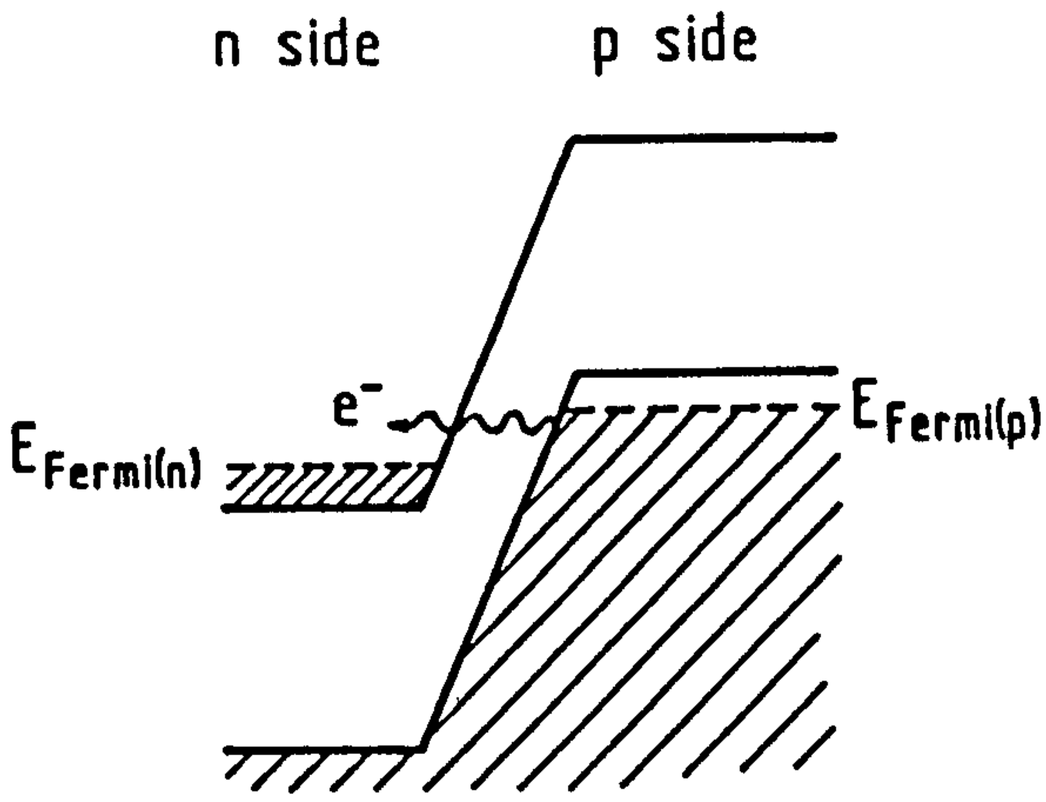


Figure 3.9: Reverse tunnelling

depletion region has a large enough electric field across it, there is a finite possibility that an n-type conduction band electron will quantum tunnel its way through the potential barrier into an empty state in the p-type valence band, resulting in a forward current. For a detector in reverse bias (the situation applicable here), the converse is true, and with p-type electrons having the same energy as empty conduction band states on the n-type side of the junction, a reverse current will flow under high enough reverse bias, as illustrated in figure 3.9.

Apart from a slight dependence via the band gap energy E_g and the diode built-in potential V_{bi}, tunneling current is roughly independent of temperature, as required. However, it is a strong function of the doping levels and reverse bias, which determine the electric field across the junction. We can adapt Thom and Yang's formula for the band-to-band tunneling current as follows :

$$\dot{N}_t = \frac{q^2 m_n^{*\frac{1}{2}} MEV}{4\pi\hbar^2 E_g^{\frac{1}{2}} A_{det}} \exp\left[\frac{-\pi m_n^{*\frac{1}{2}} E_g^{\frac{3}{2}}}{2\hbar qME}\right] \qquad e^-/\text{second/pixel} \qquad (3.33)$$

where E is the maximum electric field of a plane step junction

$= 2(V_{bi} - V)/W$ Volts/metre

M is the field factor (see below)

The field factor M gives the ratio of the average electric field in a real detector to that which would be found in an ideal plane junction. M will be ≥ 1, as the theoretical field is always exceeded due to increased fields around the detector perimeter, junction curva-

ture and material non-uniformity. M must be empirically determined for each detector design, and has been measured as being 1.3 ± 0.2 for a large variety of detectors [142]. We will treat M as a free parameter in the following fit to the low temperature dark current seen in the SBRC array.

Now we shall calculate the tunneling current for a given set of detector parameters, in order to see whether or not the actual detected dark current is consistent with this mechanism. Observed dark currents at 35 K for FPA061, a low-doped array, are approximately 100 e^-/sec/pix at both 250 and 125 mV reverse bias. For a low-doped SBRC array, the applicable doping levels are $N_A = 10^{19}\,\mathrm{cm}^{-3}$ and $N_D = 5{\times}10^{14}\,\mathrm{cm}^{-3}$. We take a reverse bias of 250 mV, a temperature of 35 K, and a detector area of $65 \times 65\,\mu$m. The junction width W is calculated to be $1.4\,\mu$m, and the electric field E as 7.1×10^5 V/m. The bandgap E_g is 0.223 eV at 35 K. This reduces equation 3.33 to

$$\dot{N}_t = -8.51 \times 10^{19}\, M \exp[-103.55/M] \qquad e^-/\text{second/pixel} \tag{3.34}$$

The number of electrons generated per second per pixel is very sensitive to the field factor M; for $M = 1.1$, $\dot{N}_t \simeq 10^{-21}\,e^-$/sec/pix, and for $M = 1.5$, $\dot{N}_t \simeq 10^{-10}\,e^-$/sec/pix — thus across the range of possible values of the fitting parameter M, the tunneling current changes by eleven orders of magnitude. However, it is immediately obvious that the predicted dark currents are far too small to account for the observed dark current. Additionally, calculating a value of $\dot{N}_t$ for a similar set of model parameters, but lowering the reverse bias to 125 mV, the dark current drops by over five orders of magnitude for any given value of M, again inconsistent with the observed approximate constancy of dark current versus reverse bias.

We must conclude that the residual dark current seen in the SBRC array is *not* due to band-to-band tunneling. The other possible mechanisms listed above, although understood theoretically, are very difficult to quantify, as they are extremely detector dependent. However, effects associated with the detector surface, and with structures surrounding the detector could easily lead to currents of the observed magnitude. The invariance of the dark current with reverse bias would indicate an effect associated with some parameter that remains fixed as a function of reverse bias, the gate voltage for example, which controls the surface potential, and thence surface leakages. The study of surface leakage mechanisms and their control via the gate voltage are discussed in detail elsewhere [121, p. 220–223].

Finally, there is the possibility that the residual dark current may in fact be due to photons, either a low level leak from outside the cryostat getting past the blanked off filter wheels that are used when making a dark current measurement, or some internal source, such as a wire, or a leak from the preamplifier. This possibility can be tested by making dark current measurements with a cold cover placed directly over the array. Such a test has not yet been performed at the UKIRT since the delivery of the science grade arrays.

3.3.3 Summary of dark current mechanisms

In astronomy, we are often dealing with far lower photon backgrounds than those typical for detectors in terrestrial applications. Thus we wish to reduce the dark current as much as possible, in order that it should not be the dominant background current. Very low dark currents ($\sim$ 10–100 e^-/second/pixel) are achievable for InSb detectors at temperatures within the range 4–77 K, that is, temperatures achievable with liquid nitrogen and helium. As discussed above, the SBRC arrays have demonstrated dark currents as low as 100 e^-/sec/pixel at 35 K, at which point G-R current has disappeared, and other effects have become dominant. By cooling further, the total dark current should continue to fall, albeit with a less steep dependence than for G-R current. Therefore further cooling should, in principle, bring further sensitivity gains. There are several reasons why this is not done however. Firstly, there is some evidence that the quantum efficiency of incorrectly processed InSb may begin to fall off steeply with temperature, and this effect should be avoided [126]. Secondly, the silicon multiplexer may begin to behave strangely at excessively low temperatures — this is the so-called 'silicon freeze-out', and staying above 30 K will reduce the chance of this happening. Both these effects have been overcome in the most recent SBRC arrays : the quantum efficiency remains high well below 30 K, and even though the multiplexer does freeze out at about 30 K, it recovers below 20 K and has been seen to function well at temperatures even as low as 3.2 K (Hoffman, private communication). The main reason for not operating below 30 K however is simply that, for devices employed in ground based imaging camera systems at least, a dark current around 200 e^-/sec/pixel will be well below the background photo-current for all but the most extremely high spatial and spectral resolutions, at short wavelengths, and therefore further cooling is unnecessary.

This will not necessarily be the case for space based cameras or ground based high resolution spectrometers however. The dark current problem needs to be re-evaluated for instances where the expected background photon flux may be on the order of 1 photon/sec/pixel. Extra cooling may prove the answer, but also, careful choice of detector semiconductor will be important. Equation 3.23 showed that the square of the intrinsic carrier concentration (n_i^2) increases exponentially as the semiconductor bandgap decreases. As we have seen, diffusion current depends on n_i^2, and G-R current on n_i, so therefore using 1–5 μm detectors in a 2 μm spectrometer, for example, will mean unnecessarily integrating excess dark current that arises due to the small bandgap that gives the 5 μm photon cutoff. Choosing a detector material with a 2.5 μm wavelength cutoff (e.g. HgCdTe) will reduce the dark current accordingly. Thus, matching the detector material precisely to the wavelength region of interest will be of greater importance in ultra low background instrumentation, in order to beat the dark current problem.

3.3.4 Photo-current

The second component of the total diode current arises when the diode is exposed to illumination. Photons with energy greater than or equal to the semiconductor bandgap ($E_\gamma \geq E_g$) hitting the diode will create an electron-hole pair, with a fractional efficiency η, known as the quantum efficiency. Diffusion will take the photo-excited charge carriers towards the p-n junction, where the junction electric field will separate them. With no external circuit attached to the diode, no current can flow; the photo-generated charge carriers collect on either side of the junction, causing a shift in the diode Fermi potentials — this voltage change is the open circuit voltage [167, p. 150]. With an external circuit, a current may flow. Under zero external bias, a current will flow — this is the short circuit current. The magnitude of the photo-current is simply :

$$I_{photon} = \eta q \phi \tag{3.35}$$

where η is the quantum efficiency

ϕ is the flux per pixel of photons with $E_\gamma \geq E_g$

The photo-current, I_{photon}, is generally regarded as a linear superposition on the dark current, I_{dark}, and independent of detector bias, V. In the convention we have adopted, the photo-current has negative sign if the positive direction is taken to be the direction

of flow of the dark current under forward bias. Hence the minus sign in equation 3.5, which we restate here :

$$I_{det}(V) = I_{dark}(V) - I_{photon}$$

It is these assumptions that the photo-current depends linearly on the photon flux, ϕ, is independent of the diode bias, V, and also that highly linear feedback resistors are used, that leads to the linearity of AC-coupled single element InSb detectors. Additionally, such detectors are typically operated near zero bias, which minimises the influence of dark current and reduces $1/f$ noise. Note however, that these assumptions are not always valid. It is known that they may break down under high flux illumination levels, and also when G-R current is a significant contributor to the overall dark current (see [121, p. 228]). However, equation 3.5 can still apply under most circumstances despite these effects, and we shall assume the equation to be valid.

3.3.5 Detector *I-V* curves

In this section, we shall present some current versus voltage (or *I-V*) curves calculated from the model developed so far. In order to derive some representative curves, we have taken the following detector parameters, known to be typical of individual InSb diodes in normally doped SBRC 62 × 58 arrays [149,142] :

$$\begin{aligned}
\text{Junction area} &= 65\,\mu\text{m} \times 65\,\mu\text{m} \\
R_{0_{diff}} &= 2 \times 10^5\,\Omega\text{/pixel (at 140 K)} \\
R_{0_{G-R}} &= 10^{10}\,\Omega\text{/pixel (at 77 K)} \\
N_A &= 10^{19}\,\text{cm}^{-3} \\
N_D &= 10^{15}\,\text{cm}^{-3} \\
\eta &= 0.7
\end{aligned}$$

Before we proceed, note again the convention that reverse bias voltages have negative sign, and that negative dark current flows for a reverse bias on the detector.

Figure 3.10 shows an *I-V* curve for an unilluminated p-n junction at a temperature of 140 K, such that diffusion current dominates. Note that, as for all unilluminated diodes, the curve passes through the origin, i.e. no dark current is generated in an unbiased diode. Also note that, after the diode is at a sufficiently large reverse bias, the

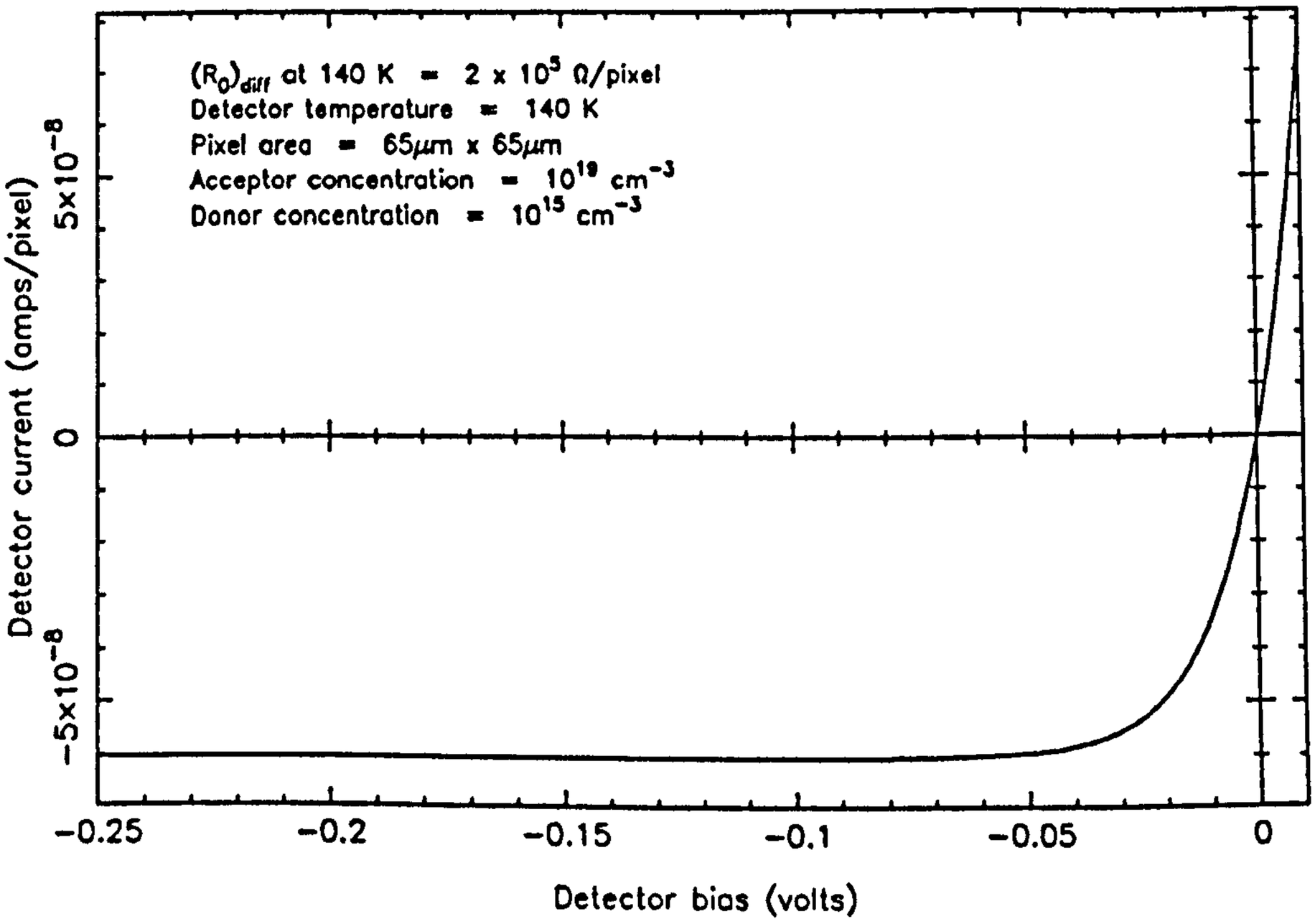


Figure 3.10: *I-V* curve for an unilluminated diffusion current dominated diode

exponential term in equation 3.6 tends to 0, such that the diffusion current saturates at $-I_{0_{diff}}$. That is, the dark current becomes independent of the reverse bias, at high temperatures, and at sufficiently high reverse bias.

Figure 3.11 shows an *I-V* curve for an unilluminated diode at 50 K, where G-R current is dominant. Again, the curve passes through the origin, but as the reverse bias is increased, the dark current does not stop increasing. This is because, after the exponential term in equation 3.8 has saturated, the $(1 - V/V_{bi})^{1/2}$ term continues to grow, as the width of the depletion region grows with reverse bias. This implies that, at temperatures between 40 and 100 K, where G-R is the dominant current source, the dark current is a function of the reverse bias and therefore changes as the diode is discharged. This is the non-linearity in the dark current mentioned at the beginning of the chapter.

Figure 3.12 illustrates the effect of illuminating the p-n junction with photons, with the detector at 50 K, where G-R current is dominant but small. The photo-current is just a linear superposition on the dark current, as discussed above, and displacing the *I-V* curve to more negative currents. As we shall see later, when in reverse bias, the dark current and photo-current act together to discharge the diode, but when discharged

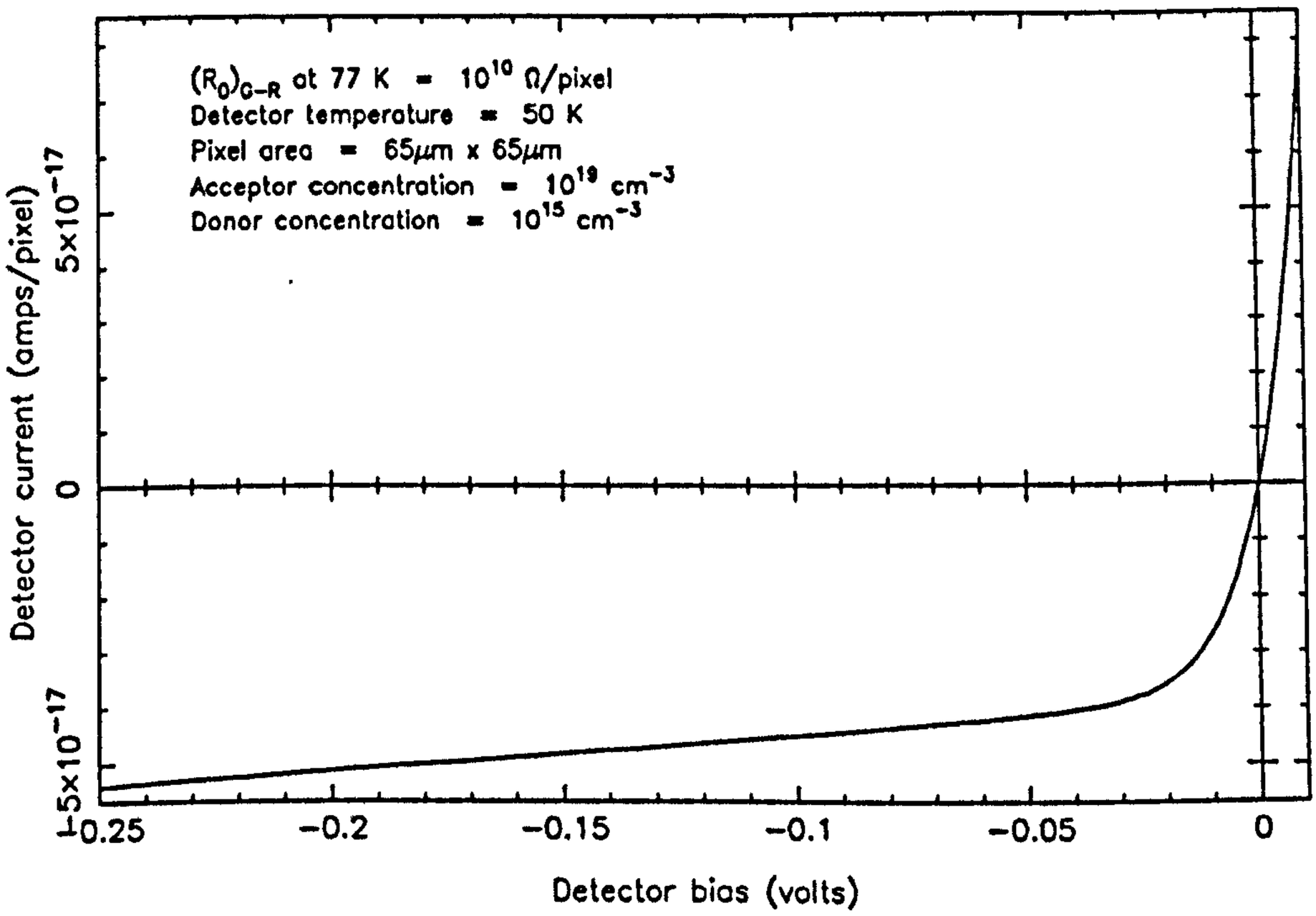


Figure 3.11: *I-V* curve for an unilluminated G-R current dominated diode

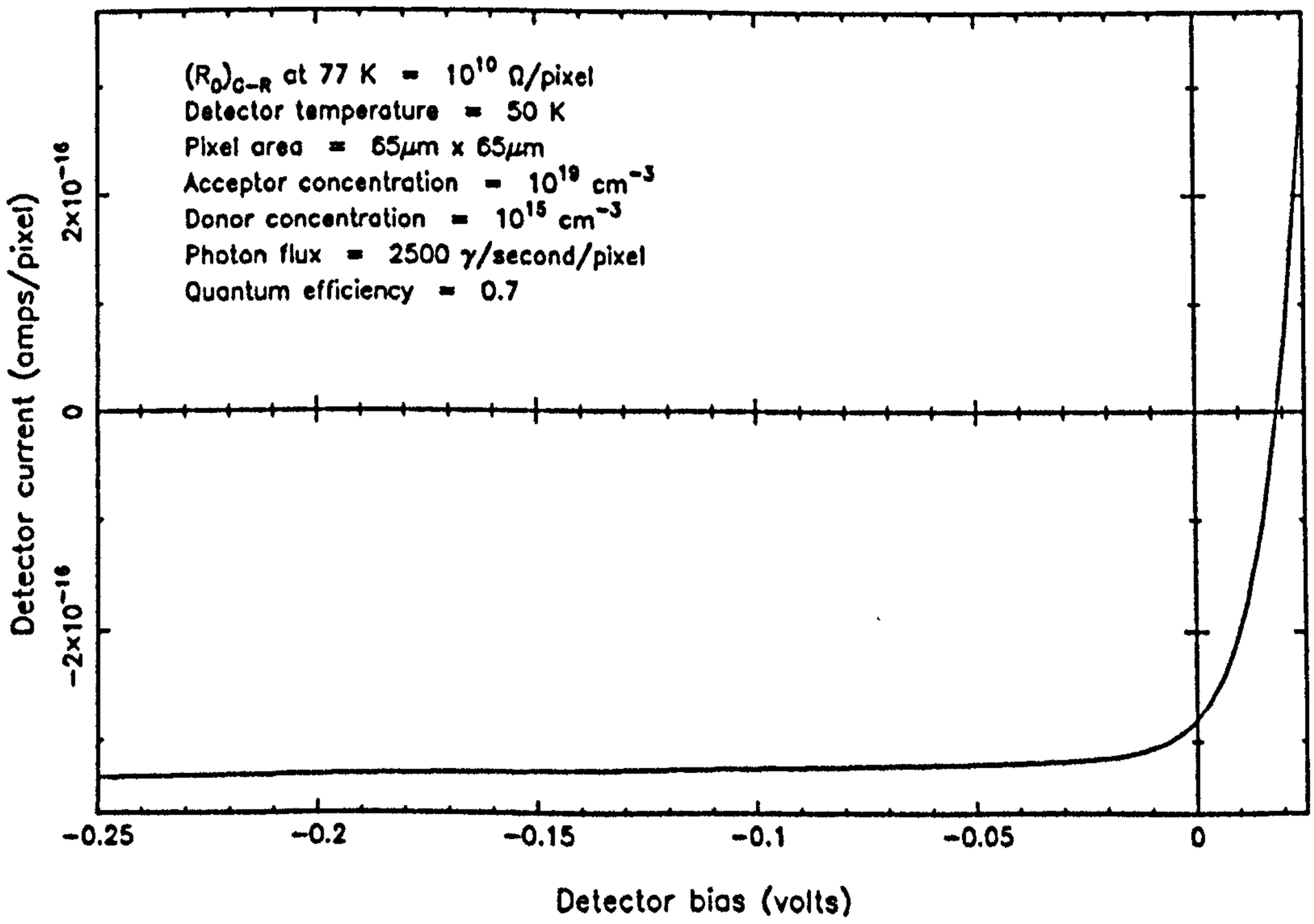


Figure 3.12: *I-V* curve for an illuminated diode

and passing through zero bias, the dark current and photo-current will begin to flow in opposite directions, ending in a balance where no net current flows. On the *I-V* curve, this is seen as the point on the positive V axis where $I = 0$, and is equivalent to the open circuit voltage. This effect is seen to be of consequence later. Finally, at zero bias, note how the *I-V* curve is displaced below the zero current axis — this is the short circuit current.

3.4 Detector capacitance

We have now dealt with the detector current, I_{det}, the top term in equation 3.4, and next we must deal with the term on the bottom, that due to the detector capacitance.

There are four main contributions to the total capacitance of the detector we outlined in section 3.2. These are due to the gate and junction capacitances of the diode itself, the capacitance of the readout MOSFET beneath, and any stray capacitances coupled into the circuit.

$$C_{total} = C_{jn} + C_{gate} + C_{FET} + C_{stray} \tag{3.36}$$

The magnitude of the detector gate capacitance, C_{gate}, depends on the length of the perimeter of the detector and the gate voltage, but is independent of the bias. Thus it is constant throughout an integration. The gate capacitance for a single SBRC array pixel has been measured to be $\sim 0.3\,\mathrm{pF}$ (Hoffman, private communication). The capacitance of the silicon MOSFET, C_{FET}, is also fixed with respect to the detector bias, and is small relative to the junction capacitance of the detector itself. The MOSFET and any stray capacitances have been found to total $\sim 0.05\,\mathrm{pF}$ (Hoffman, private communication). The junction capacitance, C_{jn}, is the largest component of the total capacitance, and as it is a function of the detector bias, it is the major source of non-linearity. We shall examine it next in detail.

3.4.1 Junction capacitance

The p-n junction acts like a parallel plate capacitor, storing charge when a voltage is placed across it. The capacitance is a function of the area of the detector, A_{det}, the width of the junction, W, and of the static dielectric constant of the material in the

junction, ϵ_s, [134] :

$$C_{jn} = \frac{\epsilon_s A_{det}}{W} \tag{3.37}$$

As the width W is a function of the reverse bias V (equation 3.10), the junction capacitance changes with the reverse bias :

$$\begin{aligned} C_{jn} &= \epsilon_s A_{det} \left[\frac{1}{\frac{2\epsilon_s}{q} V_{bi}(V_{bi} - V)\left(\frac{1}{N_A} + \frac{1}{N_D}\right)} \right]^{\frac{1}{2}} \\ &= A_{det} \left[\frac{q\epsilon_s}{2V_{bi}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)} \right]^{\frac{1}{2}} (1 - V/V_{bi})^{-\frac{1}{2}} \end{aligned} \tag{3.38}$$

Figure 3.13 shows the change in capacitance with reverse bias for a single SBRC array pixel with typical parameters. Note that it is this change in capacitance that is the primary cause of detector non-linearity in photon dominated regimes (i.e. $I_{dark} \sim 0$), and some indication of the scale of the non-linearity can be deduced from the change in capacitance seen in this figure. The other important parameter in determining the junction capacitance is the donor doping concentration N_D, although this obviously remains fixed for a given temperature. From figure 3.14, we can see that the junction capacitance drops with the donor concentration. A so-called normally doped array has $N_D = 10^{15}\,\mathrm{cm}^{-3}$, and has a junction capacitance of about 0.7 pF, whereas a low doped array has $N_D \sim 4 \times 10^{14}\,\mathrm{cm}^{-3}$, with a lower junction capacitance of 0.4 pF. The main consequence of lower doping is to reduce the junction capacitance, which in turn leads to lower read noise due to a reduced kTC component; smaller charge storage capacity as discussed below; and lower non-linearity, due to a decrease in the ratio of the bias dependent junction capacitance to the bias independent capacitance terms. Note finally that the junction capacitance is only weakly temperature dependent through V_{bi}.

3.4.2 Total capacitance

Adding the bias dependent junction capacitance to the other smaller and fixed components, C_{gate}, C_{FET}, and C_{stray}, we obtain the total diode capacitance. Recalling equation 3.4, we also need to know the partial derivative of the capacitance with respect to the bias. As only the junction capacitance is bias dependent, we differentiate equation 3.38 to obtain :

$$\frac{\partial C_{jn}}{\partial V} = \frac{A_{det}}{2} \left[\frac{q\epsilon_s}{2V_{bi}{}^3\left(\frac{1}{N_A} + \frac{1}{N_D}\right)} \right]^{\frac{1}{2}} (1 - V/V_{bi})^{-\frac{3}{2}} \tag{3.39}$$

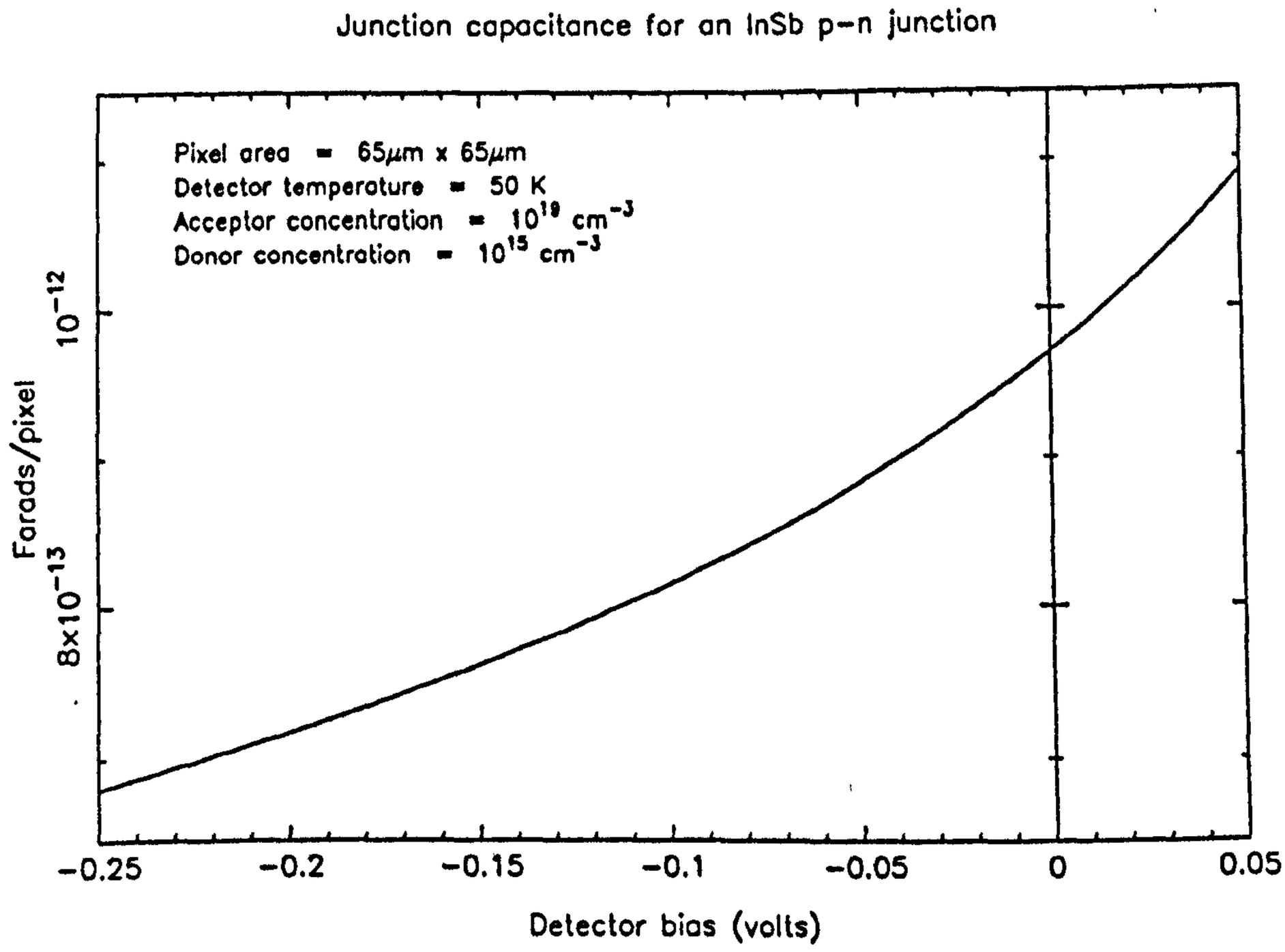


Figure 3.13: Junction capacitance as a function of bias

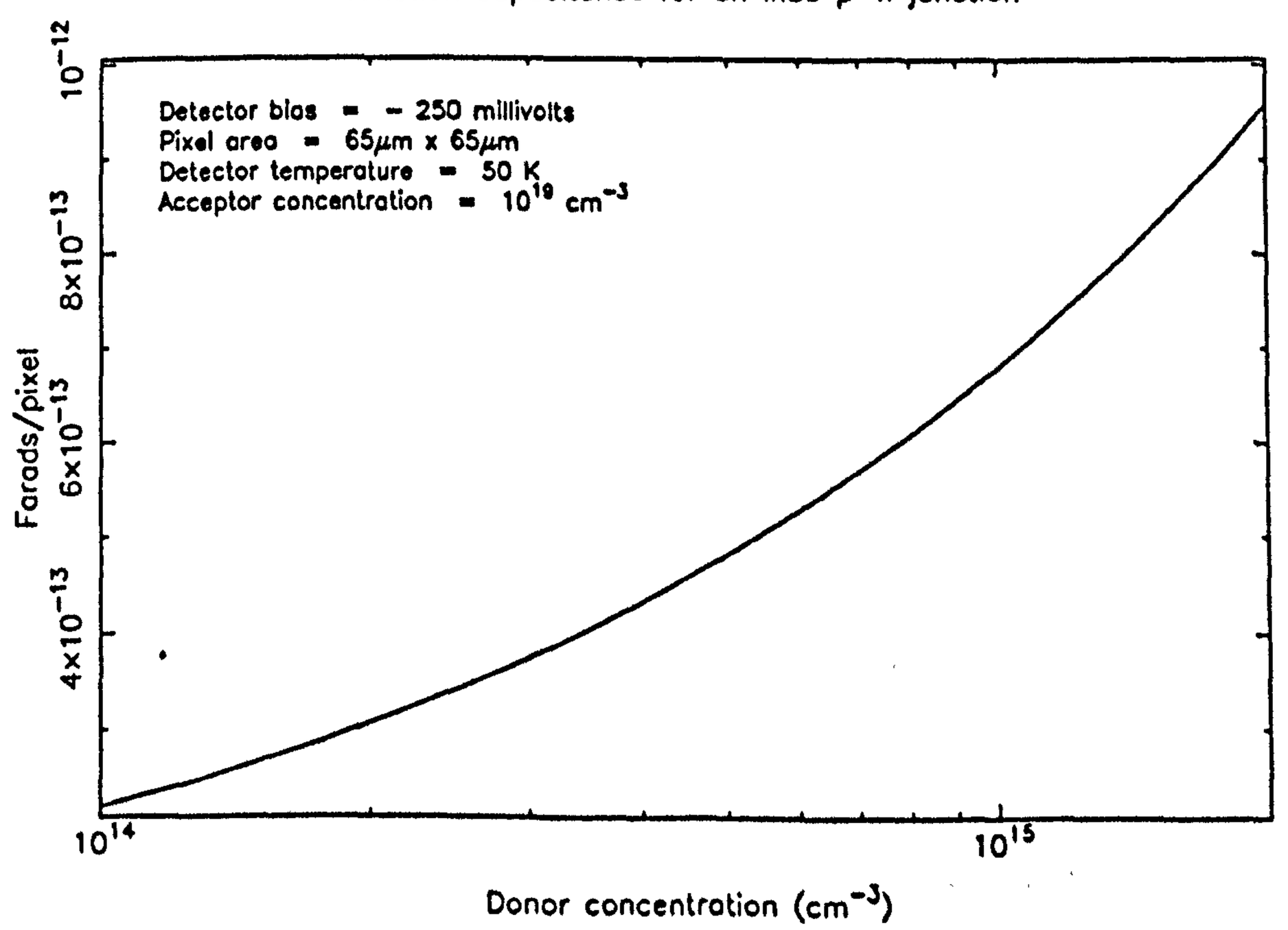


Figure 3.14: Junction capacitance as a function of doping

If we represent the fixed (bias independent) capacitances by a total capacitance C_{fix}, and the junction capacitance at zero bias by C_0, given as :

$$C_0 = A_{det}\left[\frac{q\epsilon_s}{2V_{bi}\left(\frac{1}{N_A}+\frac{1}{N_D}\right)}\right]^{\frac{1}{2}} \tag{3.40}$$

then the bottom term in equation 3.4 becomes :

$$C + V\frac{\partial C}{\partial V} = C_{fix} + C_0\left((1-V/V_{bi})^{-\frac{1}{2}} + \frac{1}{2}\frac{V}{V_{bi}}(1-V/V_{bi})^{-\frac{3}{2}}\right) \tag{3.41}$$

3.4.3 Charge storage capacity

Before proceeding, another point worth noting briefly about the capacitance of the unit cell is that it effectively determines the storage capacity of the integrating detector, known as the 'well depth'. Recalling equation 3.1, we see that the charge held on the diode Q, is proportional to the detector capacitance C, and the initial reverse bias V. During an integration period, dark and photo-generated electrons are created, and the total number of these electrons that we can 'store', i.e. the number of electrons required to discharge the reverse bias, is trivially found by :

$$\text{Well depth} = \frac{C \times V}{q} \tag{3.42}$$

where q is the electronic charge. Obviously, as we have discussed above, the capacitance of an SBRC pixel changes as the diode discharges, and this simple relation is not exactly true.

The other important parameters that help determine the well capacity are the doping levels, N_A and N_D. For high enough values of the acceptor concentration N_A, the depletion region can be considered to be an abrupt step junction, and the capacitance becomes a function of the smaller donor concentration only. Equations 3.38 and 3.20 determine this dependence, and the charge storage capacity as a function of the donor concentration is shown in figure 3.15. Note that for a normally doped SBRC array, $N_D = 10^{15}\,\text{cm}^{-3}$, whilst for a low doped array, $N_D \sim 4 \times 10^{14}\,\text{cm}^{-3}$, corresponding to charge storage capacities of $1.6 \times 10^6\,\text{e}^-$ and $1.2 \times 10^6\,\text{e}^-$ respectively, each assuming a fixed capacitance component of 0.35 pF in addition to the junction capacitance.

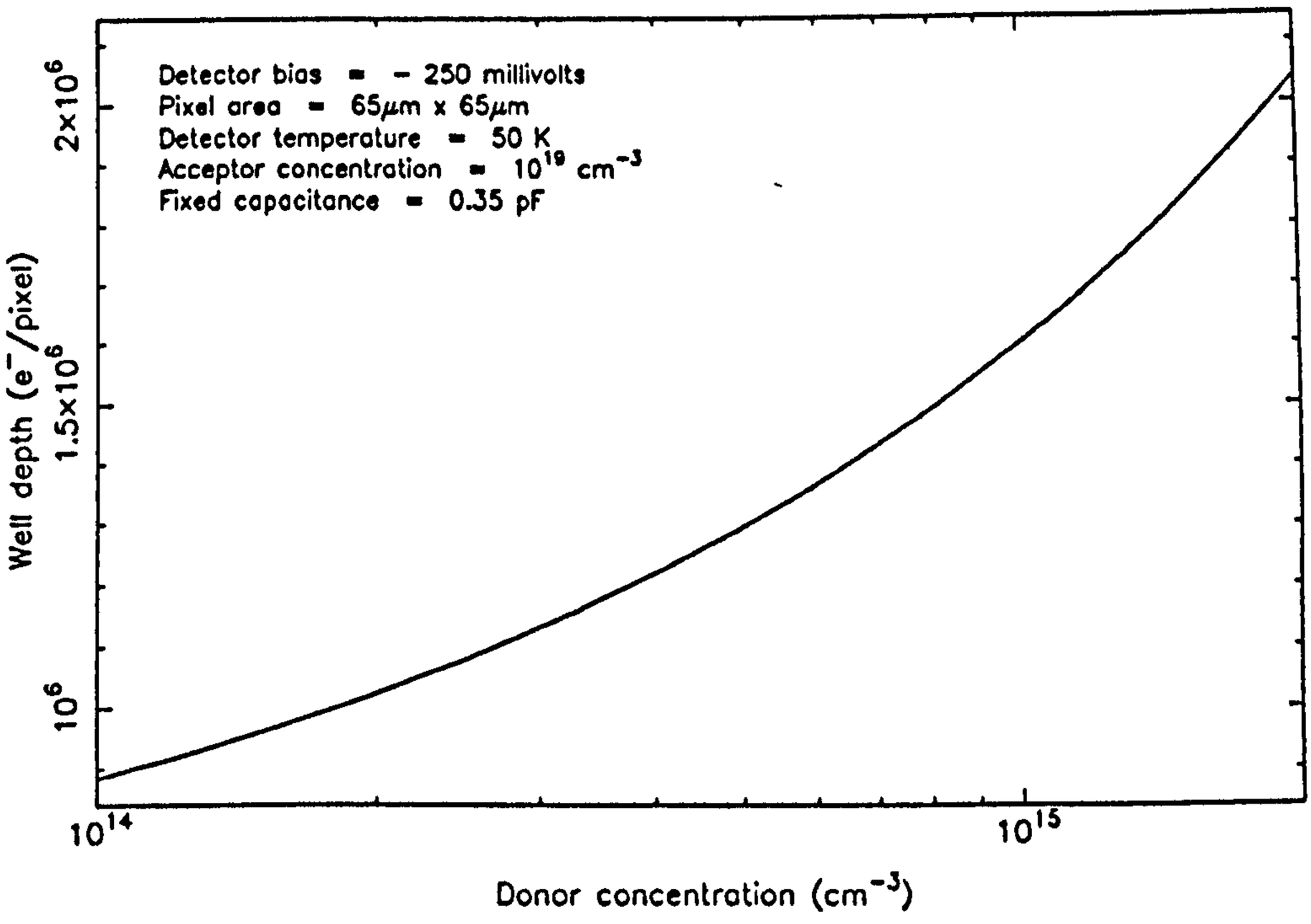


Figure 3.15: Charge storage capacity as a function of doping

3.5 Overall detector discharge equation

We are now in a position to write down an explicit form of equation 3.4, that is, the so-called 'equation of motion', that describes the voltage discharge with time, of a single unit cell in an SBRC 62 × 58 InSb + DRO array.

$$-\frac{dV}{dt} = \frac{\frac{V_t}{R_{0_{diff}}}(e^{V/V_t}-1)+\frac{2V_t}{R_{0_{G-R}}}(1-V/V_{bi})^{\frac{1}{2}}(e^{V/2V_t}-1)-\eta q\phi}{C_{fix}+C_0\left((1-V/V_{bi})^{-\frac{1}{2}}+\frac{1}{2}\frac{V}{V_{bi}}(1-V/V_{bi})^{-\frac{3}{2}}\right)} \tag{3.43}$$

As we can see, the resulting model is non-trivial, and its derivation has revealed some interesting points, which we summarise here.

- The dark current is not constant for all reverse biases. In the diffusion current dominated regime above 130 K, the dark current remains constant if the discharge is restricted, and the detector is reset before reaching (say) 40 millivolts reverse bias. In astronomical applications however, the detector temperature will always be below 77 K. Between 77 and 40 K, generation-recombination current will be

dominant. G-R current is not fixed at any point in reverse bias, due to the increasing width with reverse bias, of the depletion region in which the G-R current arises.

- Even when the detector is highly photo-current dominated, and the non-linearity of the dark current becomes unimportant, the rate of detector discharge is still not a linear function of the incoming photon flux, due to the increase in the effective capacitance of the unit cell as the reverse bias is discharged.

- There is considerable temperature dependence, particularly in the dark current terms. Even when the dark current is 'frozen out', or negligible relative to the photo-current, there remains a weak temperature dependence in the detector capacitance, via the diode built-in voltage, V_{bi}. Thus, in order to be able to hope to calibrate such devices, strict temperature control is essential, such that the various temperature dependent parameters remain as constant as possible throughout the astronomical and calibration observations.

Next we shall examine some example diode discharge curves, both in terms of the instantaneous rate of voltage discharge (dV/dt), and the diode discharge as a function of time.

3.5.1 Instantaneous diode discharge rate

Here we look at example curves, showing the instantaneous rate of voltage discharge, dV/dt, for a single SBRC array pixel, as calculated from equation 3.43. For a completely linear detector, dV/dt would be constant — that is, as the detector integrated, the rate of change of the measured signal would be constant throughout the integration, neglecting variations due to inconstant photon fluxes for example. This is not the case for the SBRC detector.

Figure 3.16 shows one extreme case, with an initial dark current 10^3 times larger than the photo-current. The model detector temperature was taken as 65 K to produce a large dark current ($\sim 3.5 \times 10^5\,e^-$/second/pixel), and although this is a very large current, unlikely to be seen in practice, it is the ratio of dark current to photo-current that is important here. One situation in which such large dark currents may occur is

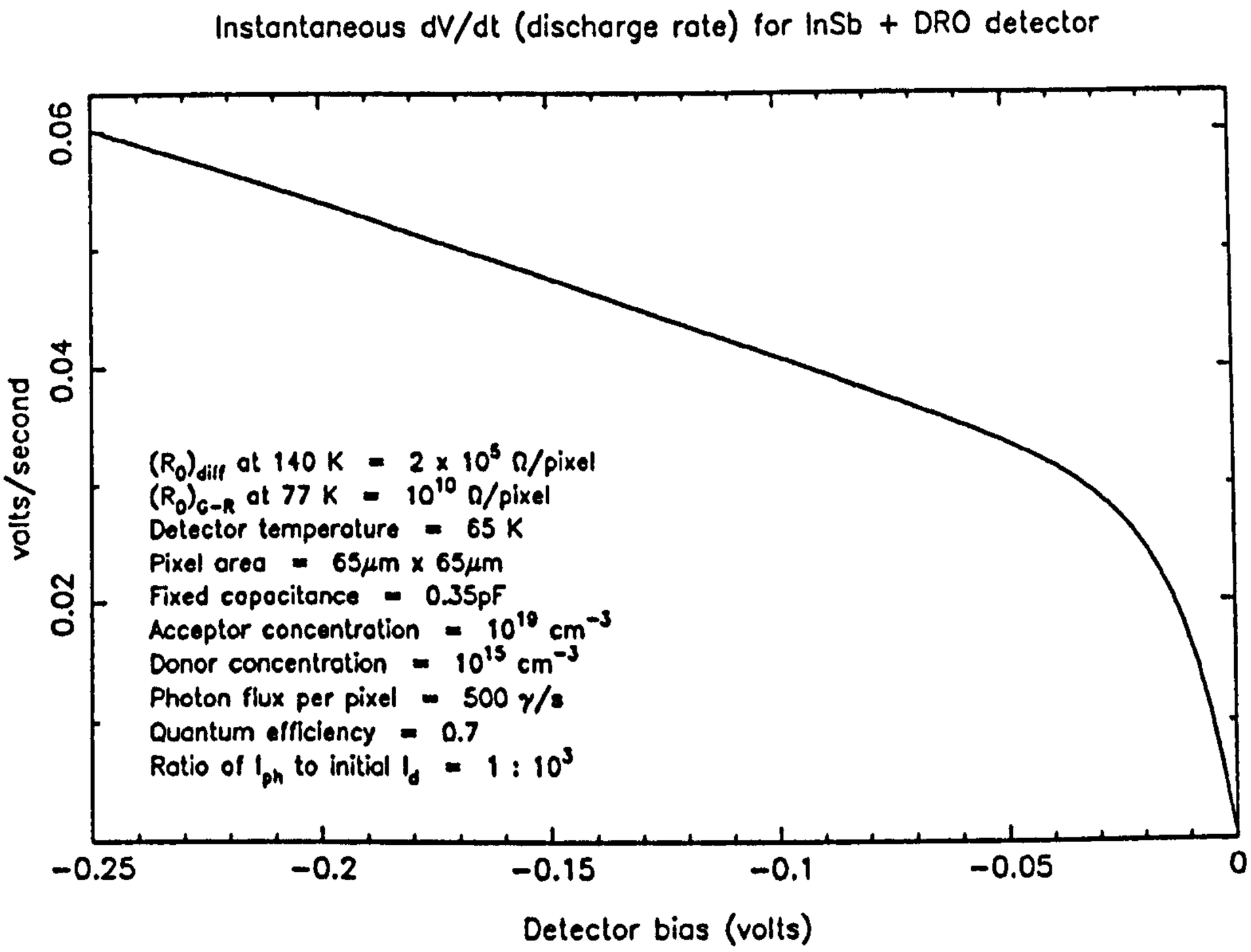


Figure 3.16: Instantaneous dV/dt — G-R dark current dominated

when IRCAM is operated in stand-by mode. That is, with no liquid helium available, the detector would operate at pumped nitrogen temperatures, providing a low sensitivity survey mode for use in conjunction with other instruments.

Note that the rate of discharge decreases by almost a factor of two between −250 and −50 mV — the $(1 - V/V_{bi})^{1/2}$ factor in the width of the depletion region comes into effect twice, as both the dark current generation rate *and* the effective conversion of current to voltage change depend on this factor. Then, below −40 mV, the rate of discharge decreases exponentially, as the dark current falls away near zero bias. At zero bias, the rate of discharge is virtually zero, with a very small residual discharge due to the small photo-current.

Figure 3.17 shows the opposite extreme, with a photo-current some 10^3 times greater than the initial dark current. This model uses a detector temperature of 50 K, giving a dark current on the order of 350 e^-/second/pixel. Even though the detector temperature will generally be lower in practice, the actual dark current will not drop much lower for the SBRC array — essentially we are using an elevated detector temperature to simulate the currents independent of temperature (surface, tunnelling) which dominate the dark current below 45 K. Either way, the curve well illustrates the photo-current dominated

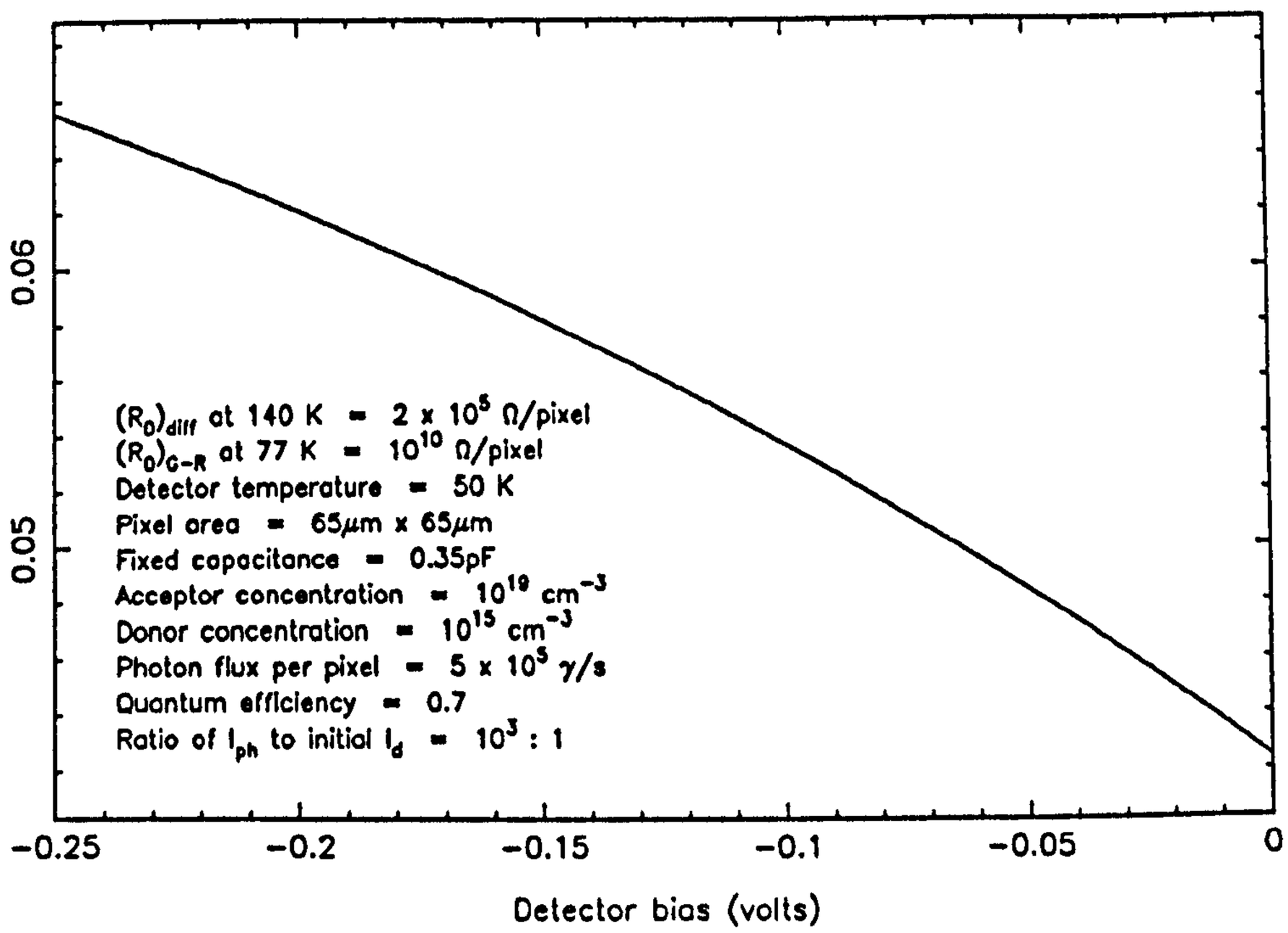


Figure 3.17: Instantaneous dV/dt — photo-current dominated

case. The value of photo-current chosen is typical of that which may be expected when imaging at K on the UKIRT, with large (2.4″) pixels.

The rate of discharge drops by a factor of only 50% in this example, between full reverse bias and zero bias, as the factor $(1 - V/V_{bi})^{1/2}$ is only seen once, due to the changing junction capacitance. The rate of discharge continues to fall passing through zero bias, but does not drop to zero. This has a consequence that we shall discuss shortly.

The final example is given in figure 3.18, and shows the intermediate case, where dark and photo-current are equal. The resulting curve is simply a combination of the two previous examples, with the almost linear decrease in rate of diode discharge between full reverse bias and −50 mV, an exponential decay below −40 mV, but not zero on passing through zero bias. Note that this combination of dark and photo-currents is fairly typical of the situation that might arise when imaging through a Fabry-Pérot étalon, and that the discharge rate ($\sim 10^{-4}$ V/s) implies a very long integration time to reach, say, half well ($\sim$ 20 minutes).

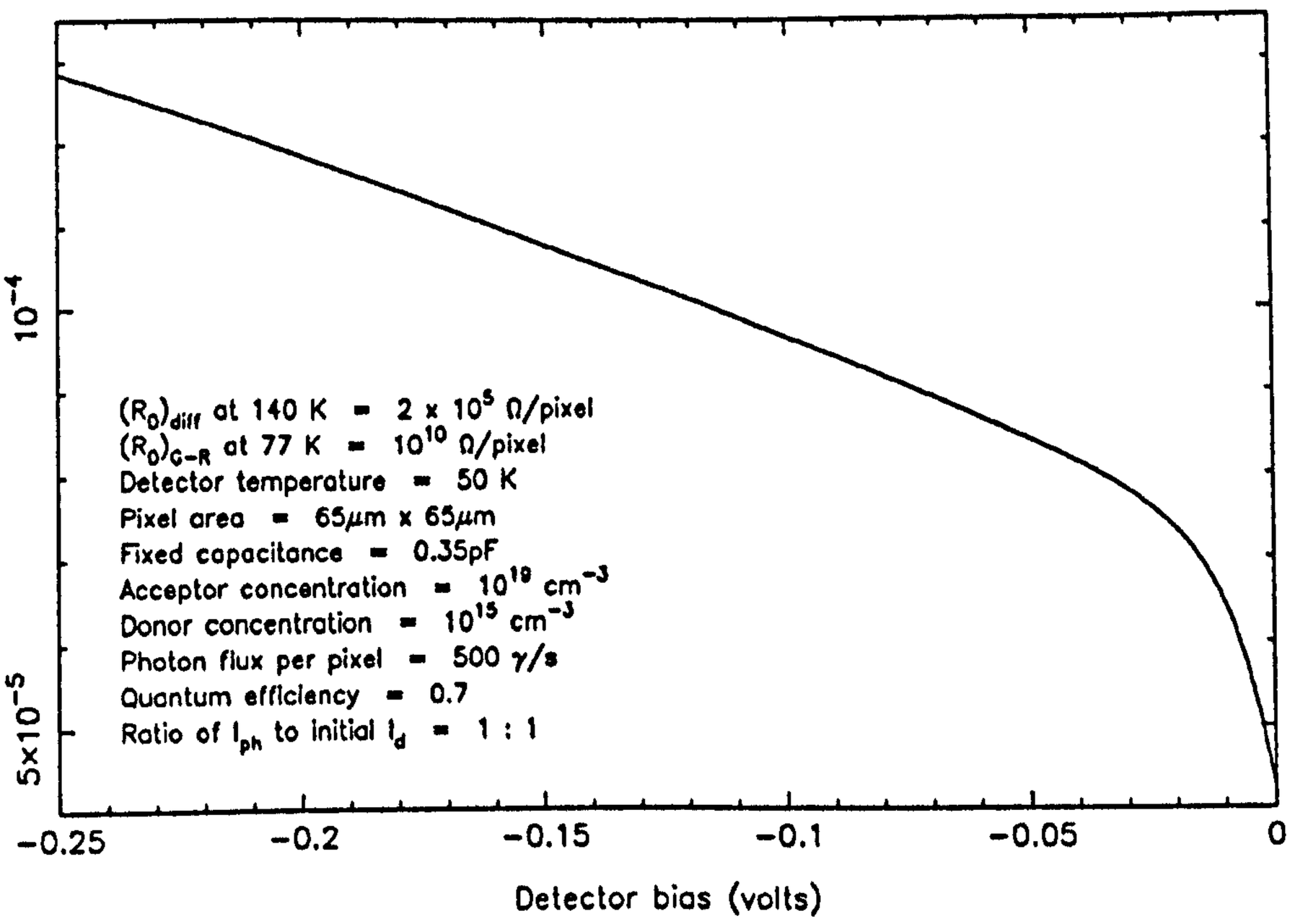


Figure 3.18: Instantaneous dV/dt — dark and photo-current equal

3.5.2 Diode discharge curves

Next we shall examine the discharge of a detector with time. For this a second order Runge-Kutta algorithm was used to numerically integrate dV/dt with respect to time. Taking the same three basic regimes outlined in the previous section (dark current dominated, photo-current dominated, and equal dark and photo-current), the following three figures show the voltage on a detector as a function of time.

Figure 3.19 shows the dark current dominated case, with the same parameters as used in figure 3.16. Again, the chosen detector temperature of 65 K means a larger than usual dark current, and thus a shorter than usual diode discharge time. But, as stated above, it is the ratio of dark to photo-current that is important here. The curve is not straight, due to the non-linearity effects, and asymptotically approaches zero bias. In the limit of zero photo-current, the diode would fully discharge after an infinite time only, if G-R or diffusion were the only source of dark current.

Figure 3.20 shows the photo-current dominated case. The parameters are typical of those likely to be encountered when imaging on the UKIRT at K with a plate scale of 2.4″/pixel, and the resulting diode discharge time of a few seconds is typical of

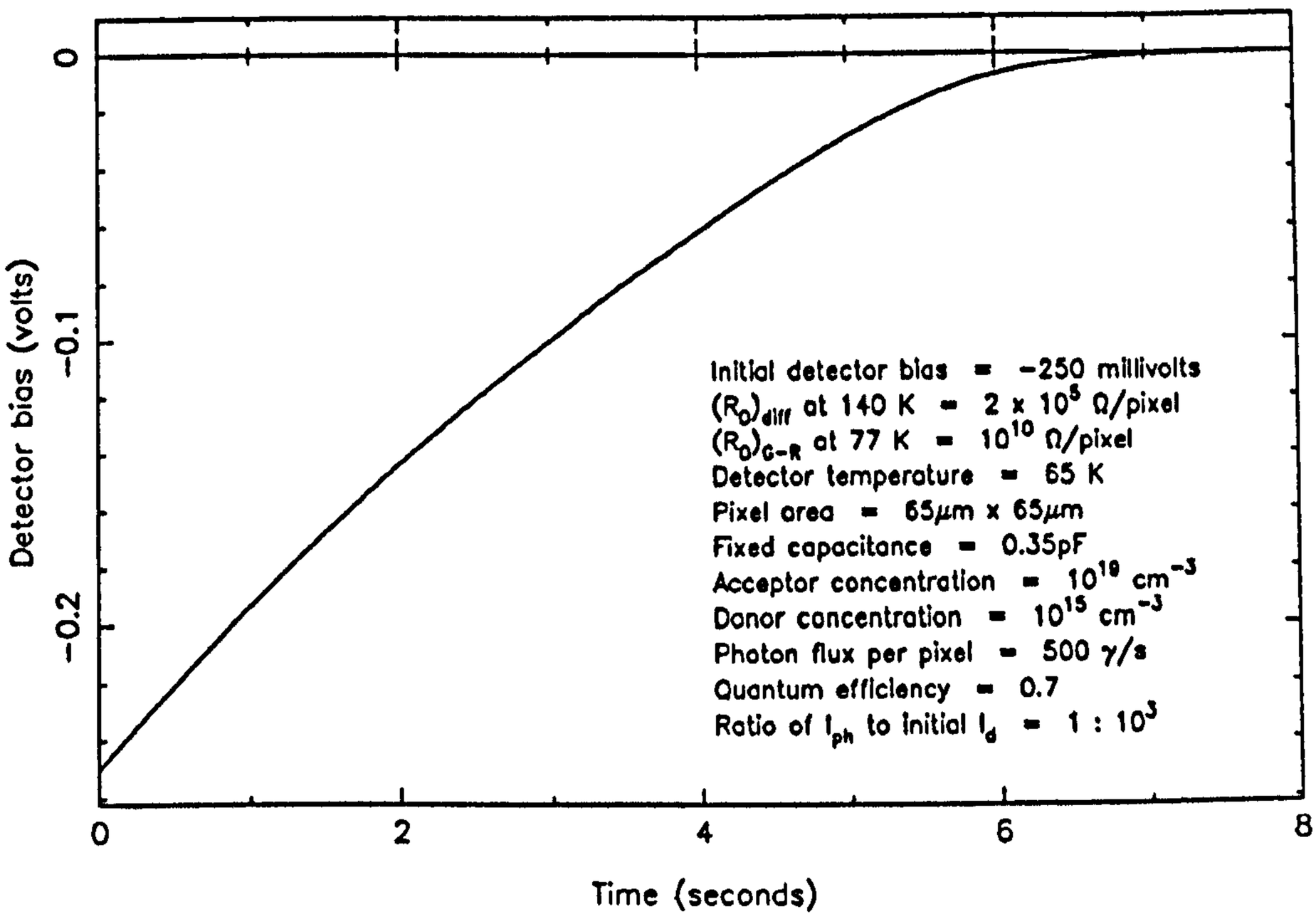


Figure 3.19: Diode discharge with time — G-R dark current dominated

integration times used under such conditions. Note how the diode discharges through zero bias, and then continues out into forward bias. The forward bias diode then generates forward dark current, i.e. in the opposite direction to that seen in reverse bias, and opposite to the photo-current. At some value of forward bias, depending upon the magnitude of the photo-current, and the temperature of the detector, the dark and photo-currents will balance, and the diode will remain fixed at this forward bias, not changing with time. Some of the consequences of this effect are discussed in the next section. Again, some qualitative measure of the non-linearity can be assessed from the deviation of the discharge curve from a straight line, the linear case.

Finally, figure 3.21 shows the interim case, with dark and photo-currents roughly equal. Again, the diode discharges non-linearly, passing through zero bias before settling at some fixed forward bias saturation value. Note that the final forward bias value is less than for the previous case, where the photo-current was 10^3 times larger. As discussed above, this interim regime is typical of Fabry-Pérot imaging, and the very long discharge time is confirmed.

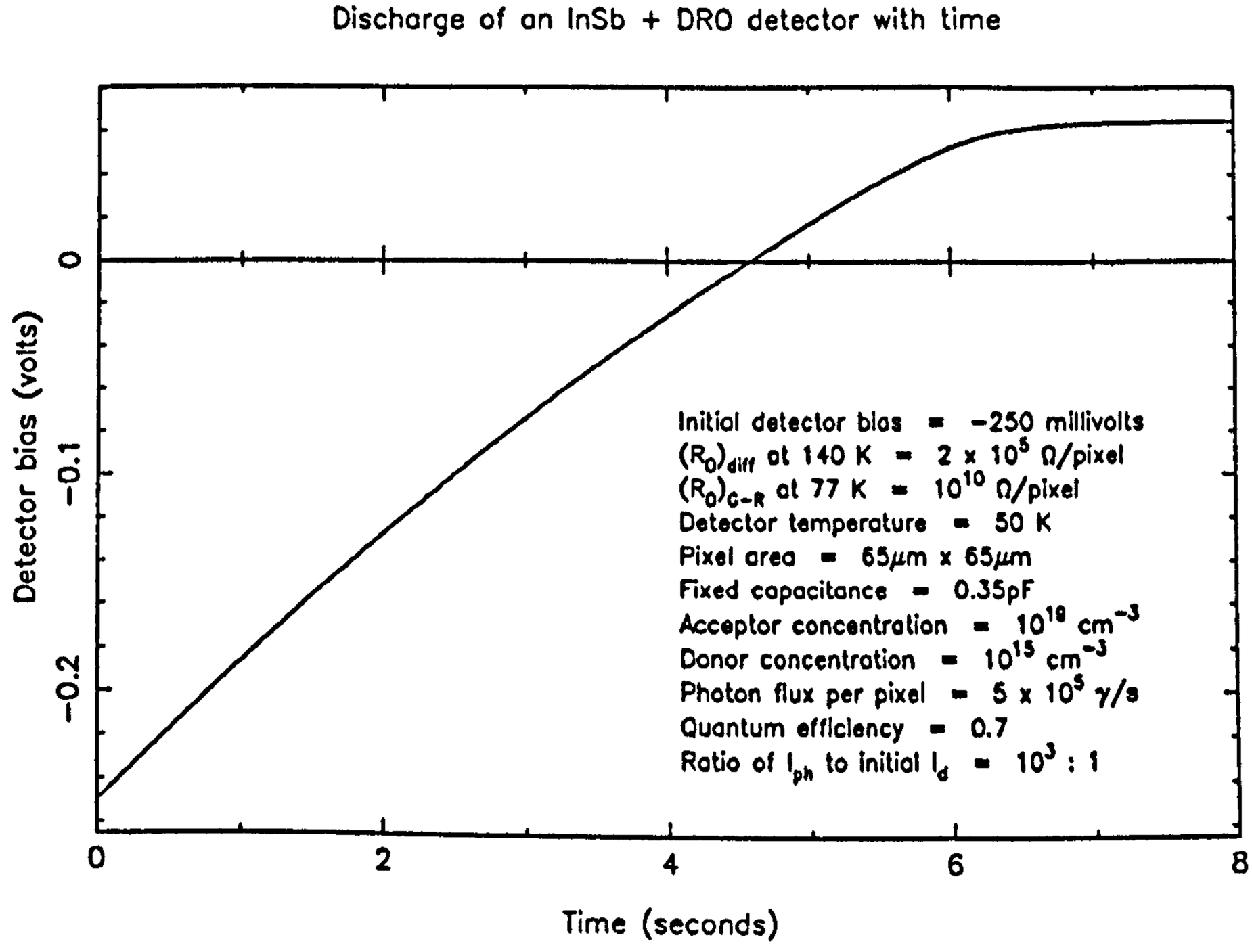


Figure 3.20: Diode discharge with time — photo-current dominated

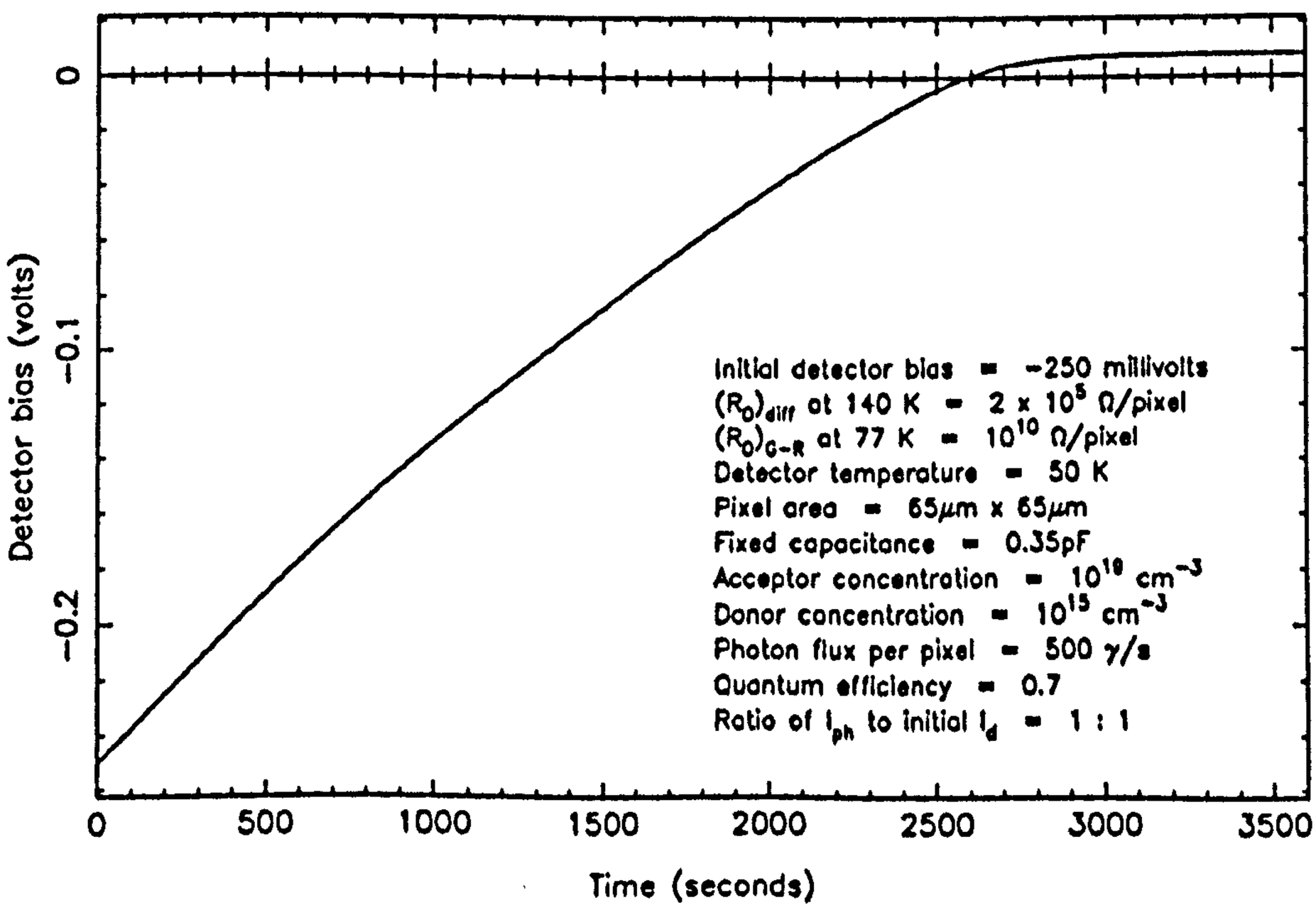


Figure 3.21: Diode discharge with time — dark and photo-current equal

3.5.3 Forward bias saturation

We have seen that a reverse biased diode illuminated with photons will not stop discharging at zero bias, but will in fact run to forward bias, where a final saturation value is reached when the forward bias gives enough forward dark current to balance the reverse photo-current. It has been suggested that leakage effects may prevent this phenomenon occurring, with the diode settling at zero bias. However, if the effect does occur, it raises the interesting possibility that some crude estimate of the source brightness may be obtained, even when the detector is saturated.

There are two basic cases where saturation due to photo-current might be encountered. Firstly, there is the case where both bright and faint objects are observed in the same field of view. Integrating long enough on-chip on the sky background to obtain background limited sensitivity on the fainter sources may well result in the brighter sources saturating. As the brighter sources are more likely to have had previous infrared photometry, they would be useful as standards against which to calibrate the faint sources.

The second case arises when the background is so bright that the array cannot be read out before saturation occurs. When photon background rates get very large, as is the case when thermal emission from the telescope and atmosphere is seen longward of $3\,\mu$m, it is possible that the detector will completely discharge before the electronics have had time to read it out. For example, with a minimum read time for the full array of 32 milliseconds, the current IRCAM system cannot handle photo-currents greater than about $3 \times 10^7\,e^-$/second/pixel, assuming a well depth of about $10^6\,e^-$. These limits are considerably exceeded when attempting broad band imaging longward of $4\,\mu$m on the UKIRT, with any reasonable pixel scale.

Thus it might be thought that no photometric information could be obtained in either of these situations. However, by letting the array saturate, and by simply non-destructively reading the voltage on each pixel many times (i.e. many repeated measurements without resetting), it is possible that some indication of the magnitude of the photo-current can be derived from the forward saturation voltage.

However, on further analysis, it appears that this capability would be of limited astronomical use. Figure 3.22 shows the final forward bias reached as a function of

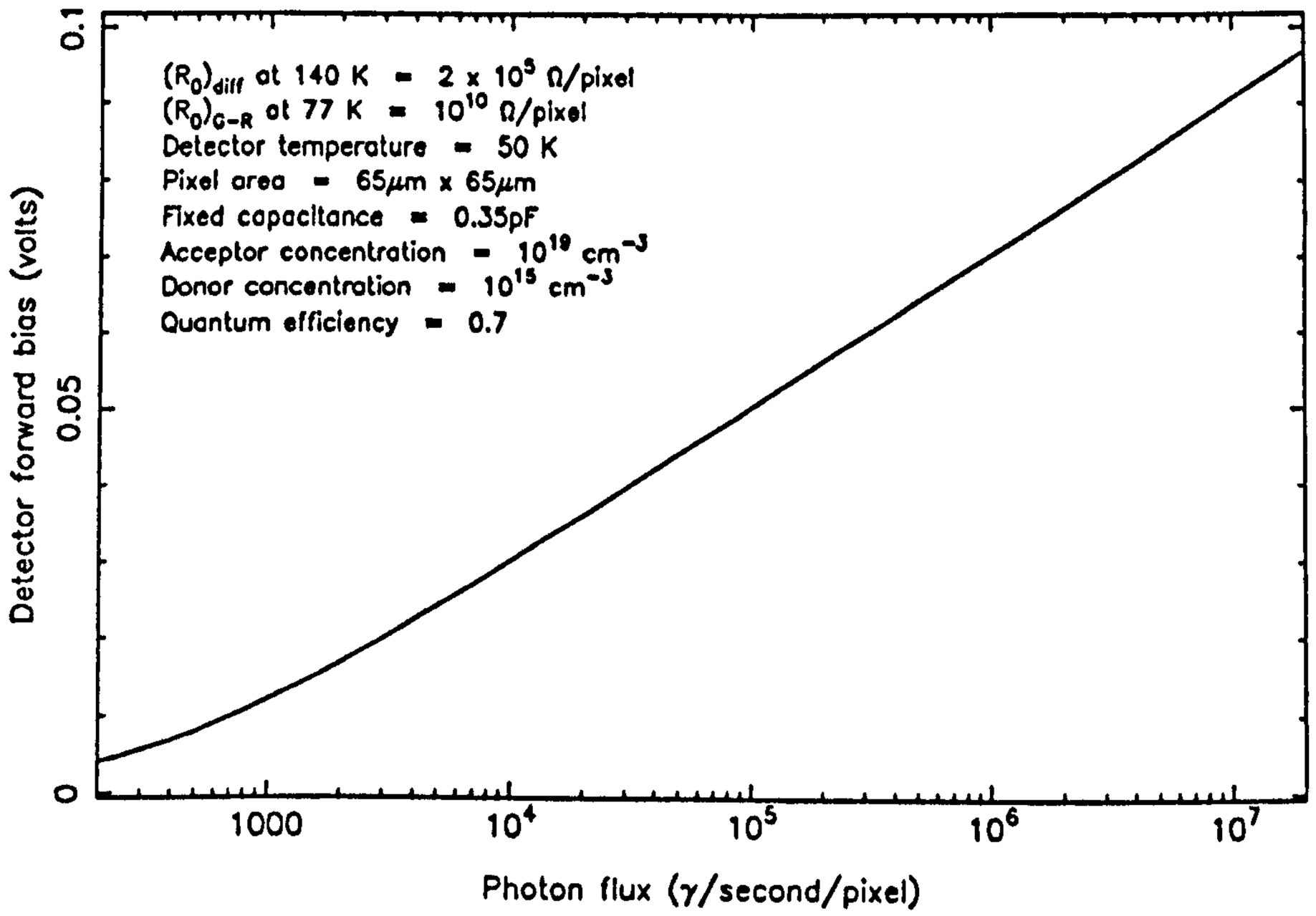


Figure 3.22: Final saturation forward bias versus photon flux

photon flux. We can see that for fluxes greater than 10^4 photons per second per pixel, the forward bias reached is proportional to the logarithm of the flux, explained by an exponential dependence of the forward dark current on the forward bias. A factor of ten increase in the photo-current brings only a 40% increase in the forward bias, and thus this method provides a very poor output signal contrast. In the situation where a bright source saturates well before the sky and fainter sources, it may prove possible to retrieve some limited photometry for the saturated brighter source. The second situation is worse for the following reason. At the thermal wavelengths where this technique would be applicable, the flux from astrophysical sources is extremely faint compared to the terrestrial background, and therefore combining this inherently low input signal contrast with the logarithmic form of the output signal contrast, it seems unlikely that this technique would be of any use.

Finally, although this technique may be of some limited use in the former situation, we have not discussed the practical considerations such as flat-fielding and linearisation in this highly non-linear region of the detector discharge curve. The difficulties that would likely be encountered there make this technique unlikely to be used in the near term.

3.6 Non-linearity

From the detector model derived in this chapter, we have seen why the SBRC array has a non-linear response to photon flux, and why the dark current is not a constant with respect to detector bias. From high thermal background imaging, to ultra-low background near-IR high resolution spectroscopy, and space borne imaging, there will be a range of situations where, say, the dark current may be dominant, or the source photons are minor perturbations on a high background, or the source photons cover a large dynamic range above both the dark current and background photo-current. All these situations may arise within the one instrument — IRCAM, for example, is capable of 3.5 μm broad band imaging, where the photon background is very large; narrow band imaging at Brγ, with the detector cooled to 35 K, at which point a large range of source photon fluxes will dominate the low photon background and dark currents; and broad band H imaging in the standby survey mode, where liquid helium is unavailable, and the detector operates at a pumped nitrogen temperature of $\sim$ 50 K, with dark current dominant.

As an introduction to a quantitative assessment of the non-linearity, we have obtained some data which show the problem graphically. Figures 3.23 and 3.24 show the output signal as a function of time from an SBRC array illuminated with a flat and constant flux. The data in the former figure were measured for a low doped SBRC array (FPA061) inside IRCAM, whilst the source of illumination was the inside of a warm cryostat cover seen through the K broad band filter. The data in the latter figure were generated using the detector model given in equation 3.43, using detector material parameters typical of a low doped array. In both figures, a linear extrapolation of the initial rate of change of output signal is compared with the actual output signal as a function of time. Whilst the units on the y-axes of the two figures are not directly comparable, the most important points are that both used an initial reverse bias of -250 millivolts, and that both discharged close to 80% of the full well. Thus they are directly comparable in terms of their deviation from linearity.

We can see that the actual data and simulated data compare well, and that the non-linearity in both cases is smooth and well behaved. Thus we can be fairly confident that the modelling gives an accurate representation of the non-linearity in an actual device, and also that it should be possible to remove this smoothly varying non-linearity either

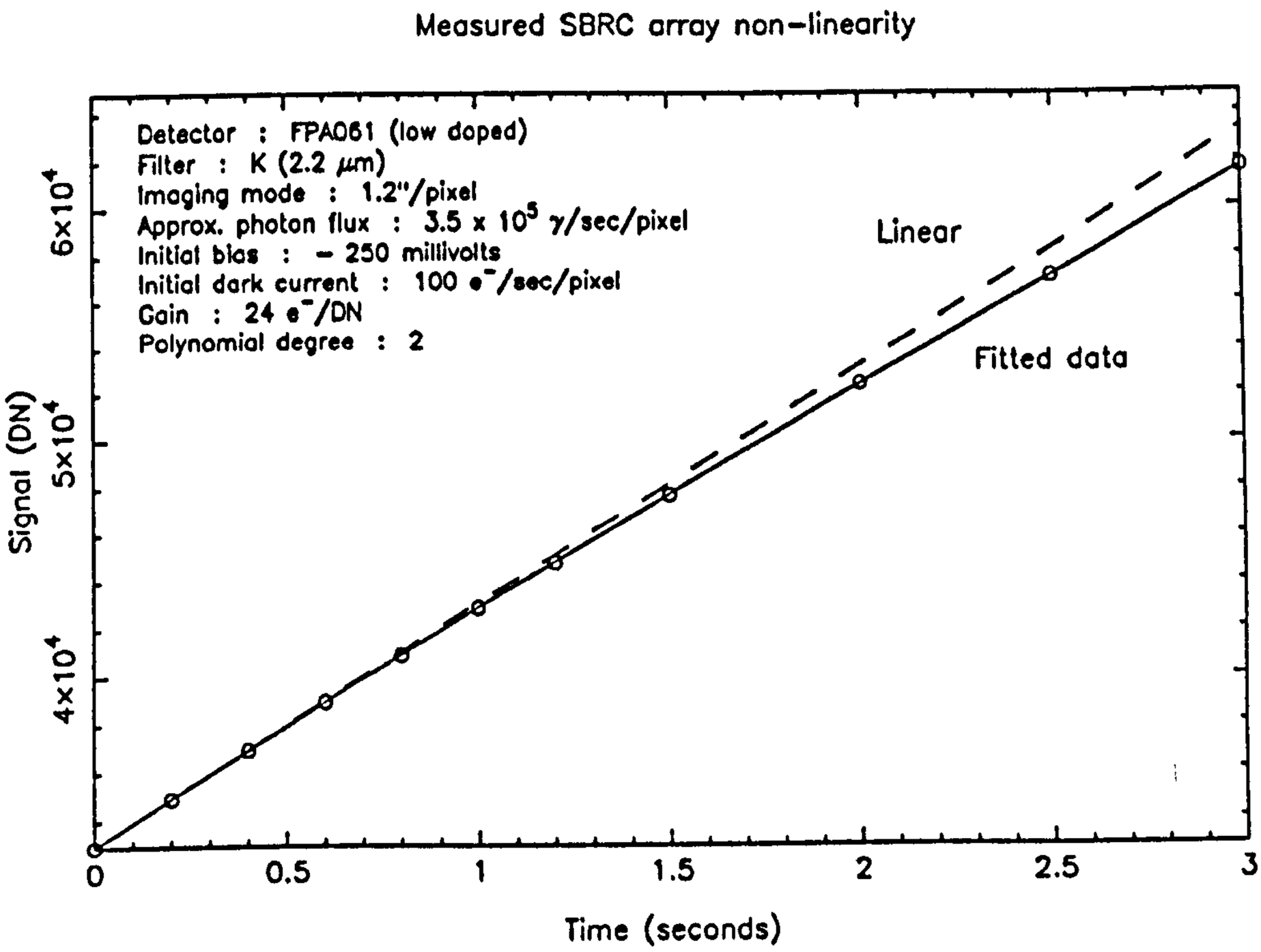


Figure 3.23: Measured non-linearity for a low doped SBRC array

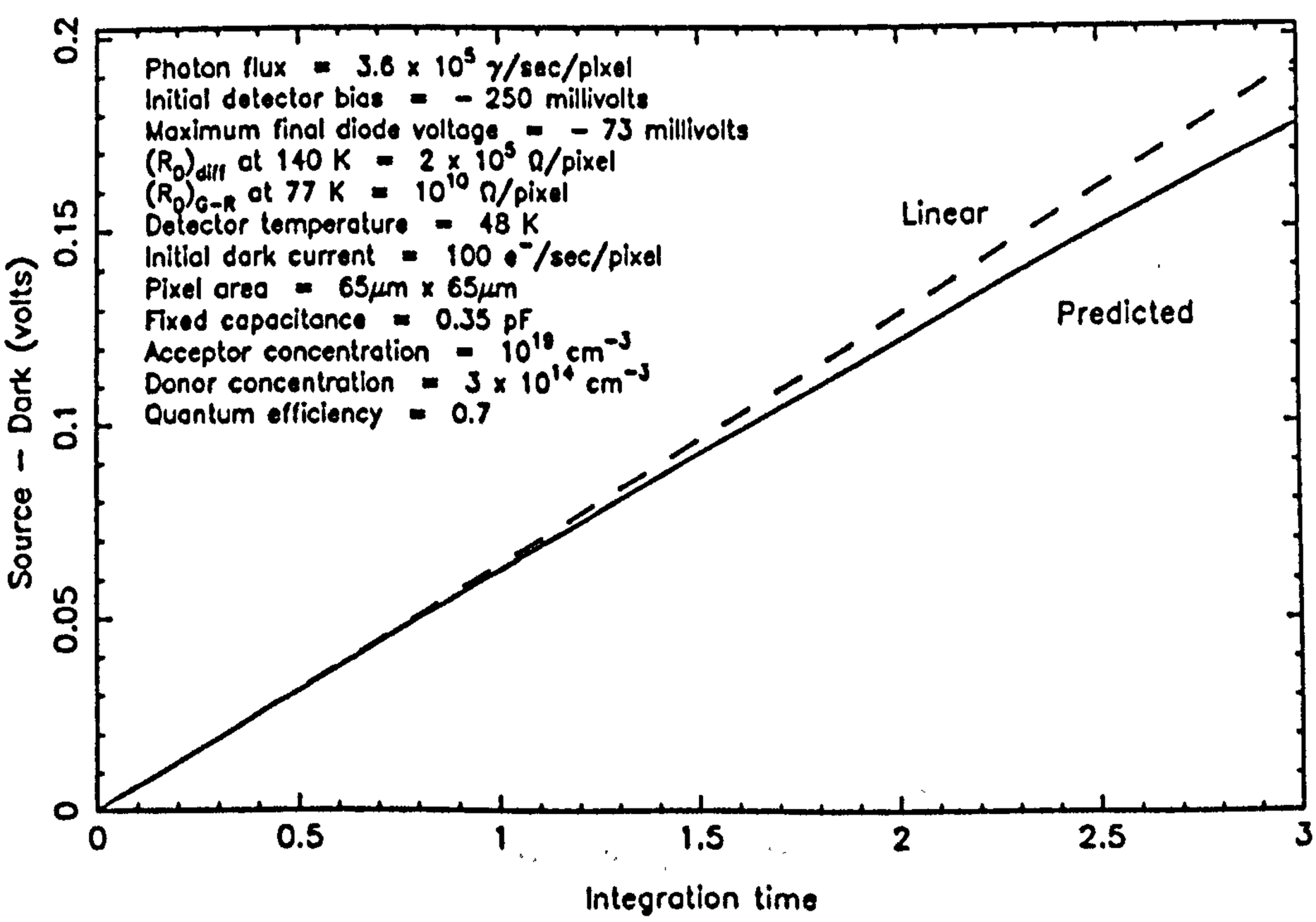


Figure 3.24: Predicted non-linearity for a low doped SBRC array

by numerical or analytical techniques.

Next we shall quantitatively evaluate the degree of non-linearity that may be expected across the wide range of likely observational conditions, before examining some algorithms and calibration techniques that may be used to reduce the non-linearity.

3.6.1 The linear case

Firstly, we need to define 'non-linearity'. We do so by addressing the linear case, then defining non-linearity as a deviation from this linear case.

In astronomy, we are concerned with finding the relative intensities of astronomical sources within an image, and across a series of images. The basic input parameter is the photon flux from the source arriving above the atmosphere for each pixel in an image. What the 'system' gives us is an electrical signal, voltage for example, for each pixel in the image. The output signal is a function of the input photon flux ϕ, the pixel quantum efficiency η (which we also use to conceal the atmospheric and optical transmission in this example), the dark current I_d, the detector and system electrical gain G, and the system electronic offset O. We define a totally linear system, in which the photo-current and dark current are independent, and where the signal S output after an integration time t is given by :

$$S = O + Gt(I_d\ /q + \eta\phi) \tag{3.44}$$

In the following analysis, we keep t fixed for all pixels, but all the other parameters may vary from pixel to pixel. In order to obtain the photon flux ϕ from the signal S for a given pixel, we proceed as follows.

Firstly, we abandon the idea of deriving the value of ϕ in absolute units — this is difficult, and in astronomy we are largely concerned with relative measurements, with one or two key absolute calibrations to zero-point the measurements. We begin by integrating for time t in the dark, i.e. with no light falling on the array, in order to measure the detector dark current. This gives us a signal S_d :

$$S_d = O + Gt(I_d\ /q) \tag{3.45}$$

Next we integrate for the same time t on a flat field, i.e. a field for which the photon

flux ϕ_f is the same for all pixels. The signal obtained is S_f :

$$S_f = O + Gt(I_d\ /q + \eta\phi_f) \tag{3.46}$$

We now wish to reduce the three 'observations', S, S_d, and S_f to give the ratio ϕ/ϕ_f, a relative photon flux, such that all pixels in the array may be compared in terms of these linear units. First we subtract the 'dark' observation from both the source and flat field frames :

$$\begin{aligned} S - S_d &= (O + Gt(I_d\ /q + \eta\phi)) - (O + Gt(I_d\ /q)) \\ &= Gt\eta\phi \end{aligned} \tag{3.47}$$

$$\begin{aligned} S_f - S_d &= (O + Gt(I_d\ /q + \eta\phi_f)) - (O + Gt(I_d\ /q)) \\ &= Gt\eta\phi_f \end{aligned} \tag{3.48}$$

Dividing, we obtain the desired ratio[1] :

$$\begin{aligned} \frac{S - S_d}{S_f - S_d} &= \frac{Gt\eta\phi}{Gt\eta\phi_f} \\ &= \phi/\phi_f \end{aligned} \tag{3.49}$$

The image is now calibrated in relative photon flux units, with all pixel to pixel variations nominally removed. The final step is to remove the background photon flux ϕ_b, due to non-astronomical thermal and non-thermal emission sources in the beam. After flat fielding, the background flux over the given image is determined to be ϕ_b/ϕ_f, and then the final relative astronomical intensity value in a given pixel ϕ' is :

$$\phi' = (\phi/\phi_f) - (\phi_b/\phi_f) \tag{3.50}$$

The values of ϕ' across the image should now tell us the relative intensities of the astronomical sources within that image, if the conditions of linearity and parameter independence assumed are true.

Thus, if two objects in the image have absolute extra-atmospheric intensities in the ratio 2:1, then their calculated ϕ' values should be in the same ratio. If in fact, after the reduction process outlined above, the ratio of the two values of ϕ' is (for example)

[1]Note that at this stage we would usually rescale the resulting image, multiplying it by the median value of a fixed region of the $S_f - S_d$ image in order to allow comparison between images calibrated with different flat field images — we omit this rescaling step here for clarity, its omission not affecting the final result

1.7:1, then we could say the 'system' was 30% non-linear. We can more formally say : For a system measuring two input photon fluxes ϕ_1 and ϕ_2, and outputting the values ϕ_1' and ϕ_2' after reduction, we define the non-linearity Γ as :

$$\Gamma = 1 - \left[\left(\frac{\phi_2' - \phi_1'}{\phi_1'} \right) \Big/ \left(\frac{\phi_2 - \phi_1}{\phi_1} \right) \right] \tag{3.51}$$

This figure, Γ, is measure of the practical non-linearity of an astronomical system, i.e. it is the deviation from linearity *after* corrections for system offset, system gain, dark current, and quantum efficiency have been made on a per pixel basis, by a method that assumes a linear and independent system. It is a normalised non-linearity that is undefined when $\phi_1 = \phi_2$, as opposed to the more usual definition, where this special case gives a non-linearity of zero. The definition presented here has the quality that for a non-linear system it reduces to zero nowhere, but will reduce to zero for a linear system. For the rest of this section, unless otherwise stated, any figure given relating to the non-linearity of a system will be taken to imply the parameter Γ as defined in equation 3.51.

3.6.2 Non-linearity of the InSb + DRO array

We have seen in this chapter that the SBRC InSb + DRO array is inherently non-linear in its response to photon illumination due to the dependence of the circuit capacitance on the discharging reverse bias. High photon fluxes give smaller output signals than would be anticipated by a linear scaling of the signal output by pixels illuminated by lower photon fluxes. The dependence of the dark current on the reverse bias complicates the situation further — the integrated dark current contribution to the total output signal of a pixel under high illumination will be less than for a pixel under low illumination, as the high flux will more rapidly discharge the reverse bias, thus making I_{G-R} decrease more quickly.

Thus, the assumptions made in deriving the linear reduction procedure in the previous section are not true for the SBRC array, and we would expect the non-linearity parameter, Γ, to be non-zero if that method were followed.

We shall now examine the degree of non-linearity, Γ, as a function of the major parameters input to the model of the SBRC array given in equation 3.43. There are five effective regimes of operation which we shall concentrate on :

- Regime 1 — high background photon flux, where dark current is negligible, and the photon flux from astronomical sources is only a minor perturbation on the background (e.g. thermal imaging, $\lambda \geq 3\,\mu$m)

- Regime 2 — medium background photon flux, dark current again negligible, but with astronomical source fluxes ranging from a small fraction of the background to much greater than the background (e.g. broad-band imaging, $\lambda \leq 2.5\,\mu$m)

- Regime 3 — low photon background flux, smaller than the dark current, and astronomical source fluxes ranging from a fraction of the dark current to many times the dark current (e.g. high spectral resolution imaging, $\lambda \leq 2.5\,\mu$m)

- Regime 4 — low photon background flux, smaller than the dark current, and with astronomical source fluxes small compared to the dark current (e.g. high resolution faint object spectroscopy, $\lambda \leq 2.5\,\mu$m)

- Regime 5 — interim case, with background flux, dark current, and astronomical source flux all roughly equal (this could arise at a transition between any of the other four regimes)

We have graphed the non-linearity parameter Γ as a function of input photon flux for each of these regimes. The figures are derived as follows. The relevant model detector parameters are input, including initial reverse bias, detector temperature, R_0s for both G-R and diffusion current, acceptor and donor concentrations, pixel area, fixed capacitance, quantum efficiency, and on-chip integration time. A dark current exposure is simulated. Then a background photon flux is input, and a background + dark current exposure simulated. Lastly, a range of source photon fluxes is chosen, and for each one, a source + background + dark current exposure is made, using the same integration time throughout. The data are 'reduced' according to the method outlined in section 3.6.1, and the non-linearity Γ calculated according equation 3.51.

In each case, the input model parameters are given in the figure. Two numbers of interest arising from the model are the initial dark current, i.e. the dark current at the starting reverse bias, and the maximum final diode voltage. The latter number is a measure of how close to discharged the detector came in the specified integration time in the case of dark + background + the largest requested source flux. The detector was not allowed to saturate for any given flux. Throughout, the integration time was

chosen such that the minium reverse bias any pixel reached was between −50 and −70 millivolts, starting at an initial reverse bias of −250 millivolts. In this way, 80% of the available dynamic range was used, sufficient numbers of charge carriers were integrated to overcome read noise for at least the brightest sources, and potential problems occurring near zero bias were avoided. Under actual observing conditions, some sources might be allowed to saturate in order to integrate for longer on faint sources, and on the background. This does not affect the general results presented in the non-linearity curves however, as any given combination of background, source flux, and dark current may be found by looking at the various curves.

It is worth noting again that the actual magnitudes of the dark current, background photon flux, and range of source fluxes are not important; it is the ratios of these components that is important. Thus the integration time used is merely given as a guide, and this number may scaled according to the absolute magnitude of the currents.

In figure 3.25, we have simulated Regime 1, where the dark current is negligible, the background high, and the source fluxes only a small perturbation on the background. Thus the dynamic range in the output signals is small, with all source signals sitting on a large pedestal due to the background — the non-linearity is about 0.2%, for a source photon flux 100 times smaller than the background. Similarly, in figure 3.28 a maximum non-linearity of about 0.2% is found in in Regime 4, where dark current is dominant, the background is low, and the sources are faint.

Contrast this with the situation where the range of source photon fluxes is large, the background low, and the dark current negligible, which we have defined as Regime 2 — the results are shown in figure 3.26. The dynamic range in the output signals is large, and so is the non-linearity. A non-linearity of 10% is seen in this regime, typical of broad-band imaging observations. Note that an on-chip integration time of 1.5 seconds is used, in order not to saturate on the brightest source. However, the background is low, and would require an on-chip integration time around 20 seconds in order to be shot noise limited, as opposed to read noise limited. So, the array may be read out more rapidly to allow (for example) wide dynamic range stellar photometry, or alternatively, the background might be integrated up to overcome the read noise, in order to provide the maximum sensitivity to very faint sources near the background. In the former case, a wide dynamic range is obtained at the expense of linearity, and the latter case gains

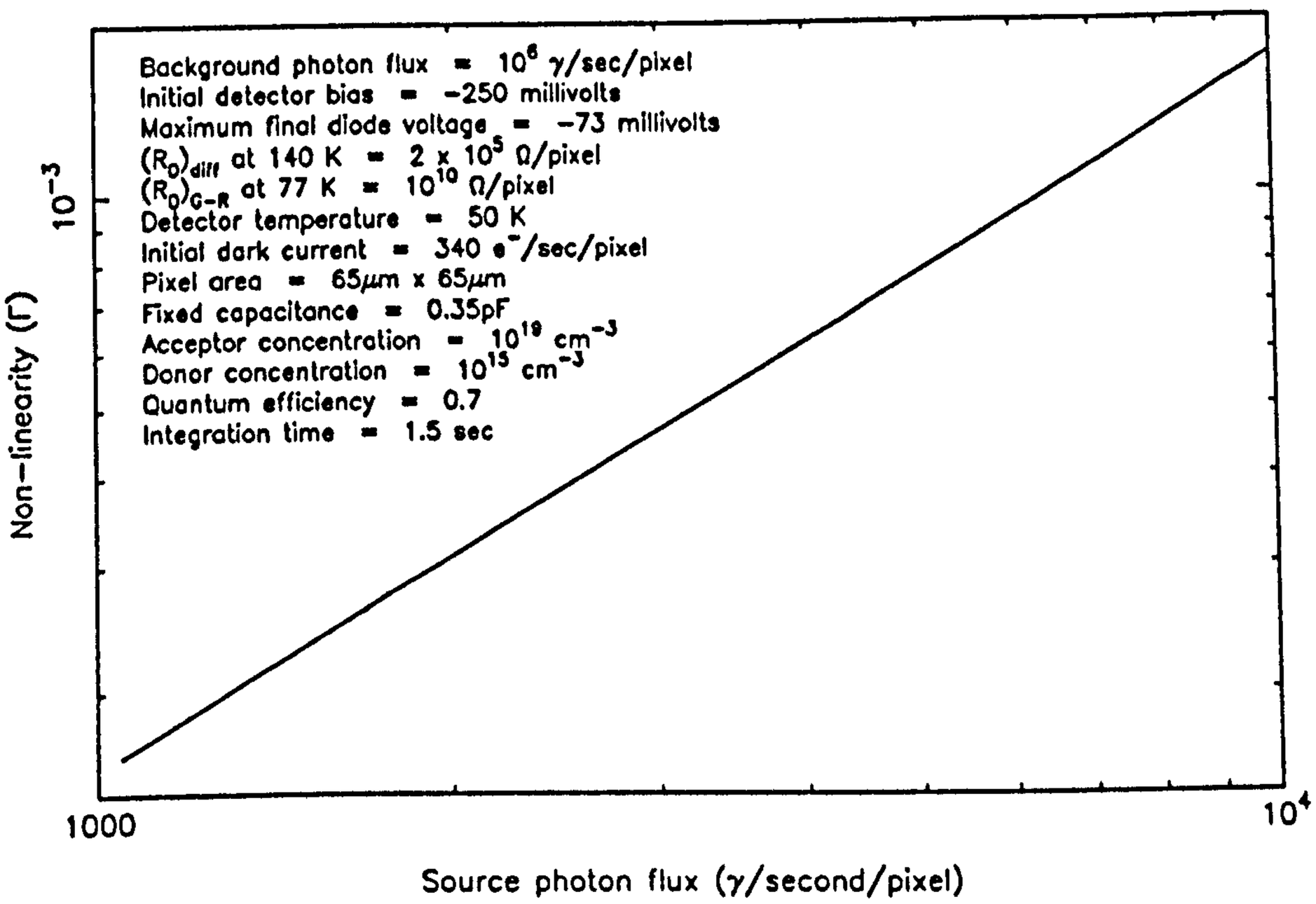


Figure 3.25: Non-linearity versus source flux — Regime 1

faint object sensitivity and linearity at the expense of dynamic range.

Figure 3.27 shows the non-linearity in Regime 3, where the background is low, the dark current larger, but with a wide range of source fluxes again dominating the discharge. This is essentially analogous to Regime 2, and again, the maximum non-linearity is about 10%.

Finally, figure 3.29 shows the non-linearity obtained in Regime 5, where the dark current, background and source photo-currents, all have roughly the same magnitude, as may arise at a boundary between any of the other regimes. This mixed situation gives a maximum non-linearity of about 6%.

To summarise :

- When the range of source photon fluxes gives only a small perturbation on either the dark current or the background photo-current (Regimes 1 and 4), the non-linearity is small, and typically less than 1%.

- When the brightest source fluxes are much greater than the dark current or background flux, and the output signal dynamic range is large (Regimes 2 and 3), so

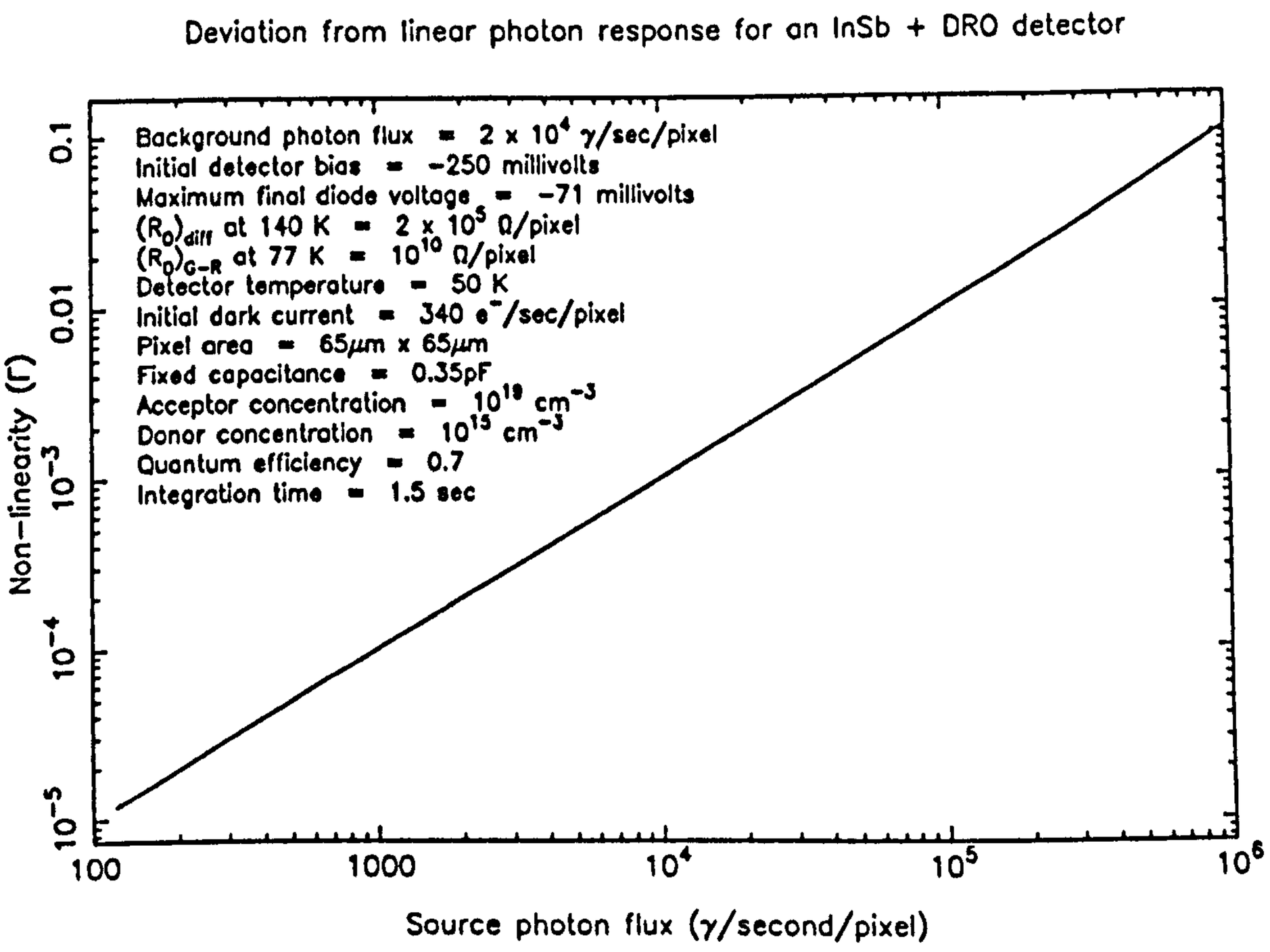


Figure 3.26: Non-linearity versus source flux — Regime 2

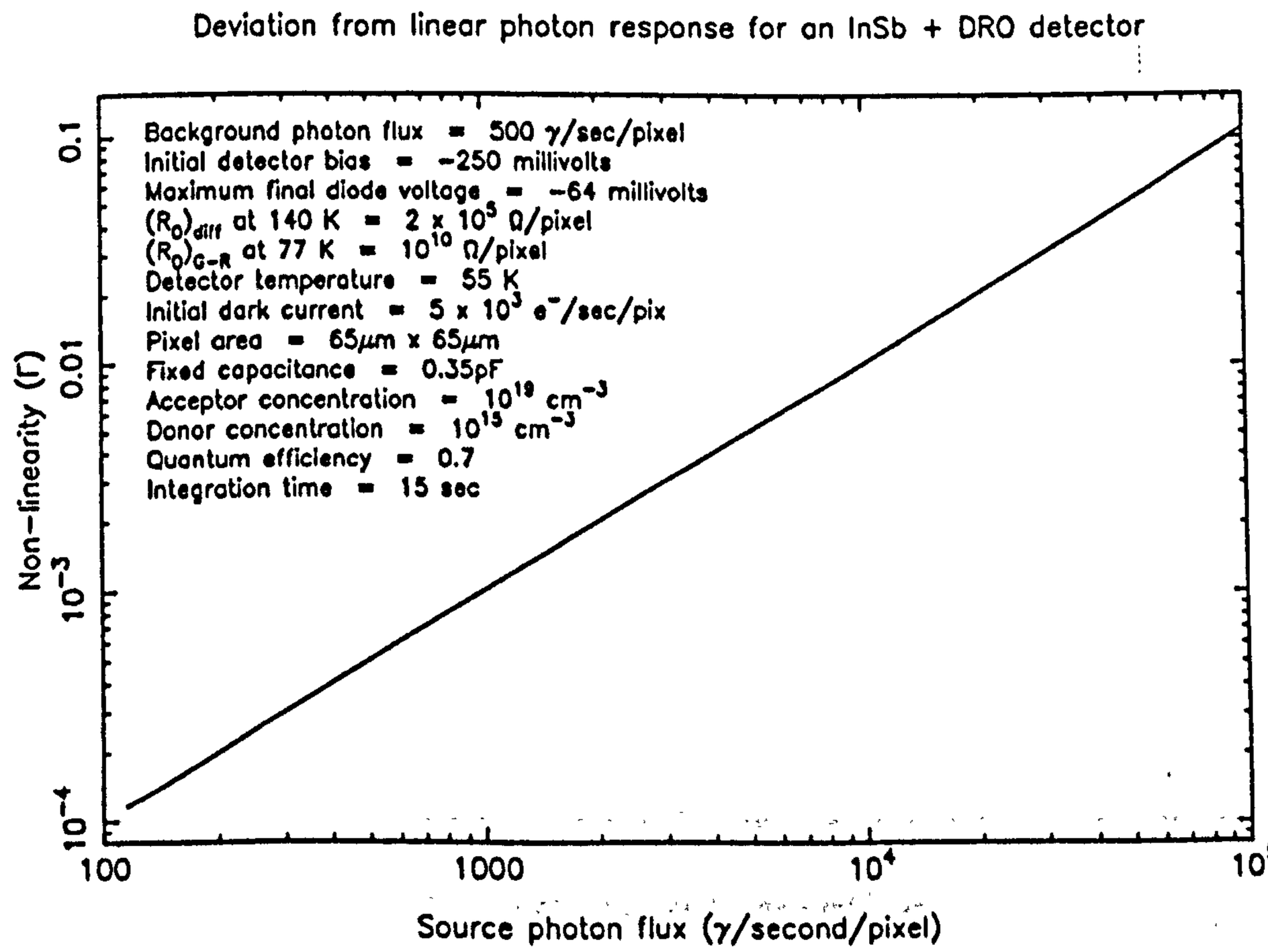


Figure 3.27: Non-linearity versus source flux — Regime 3

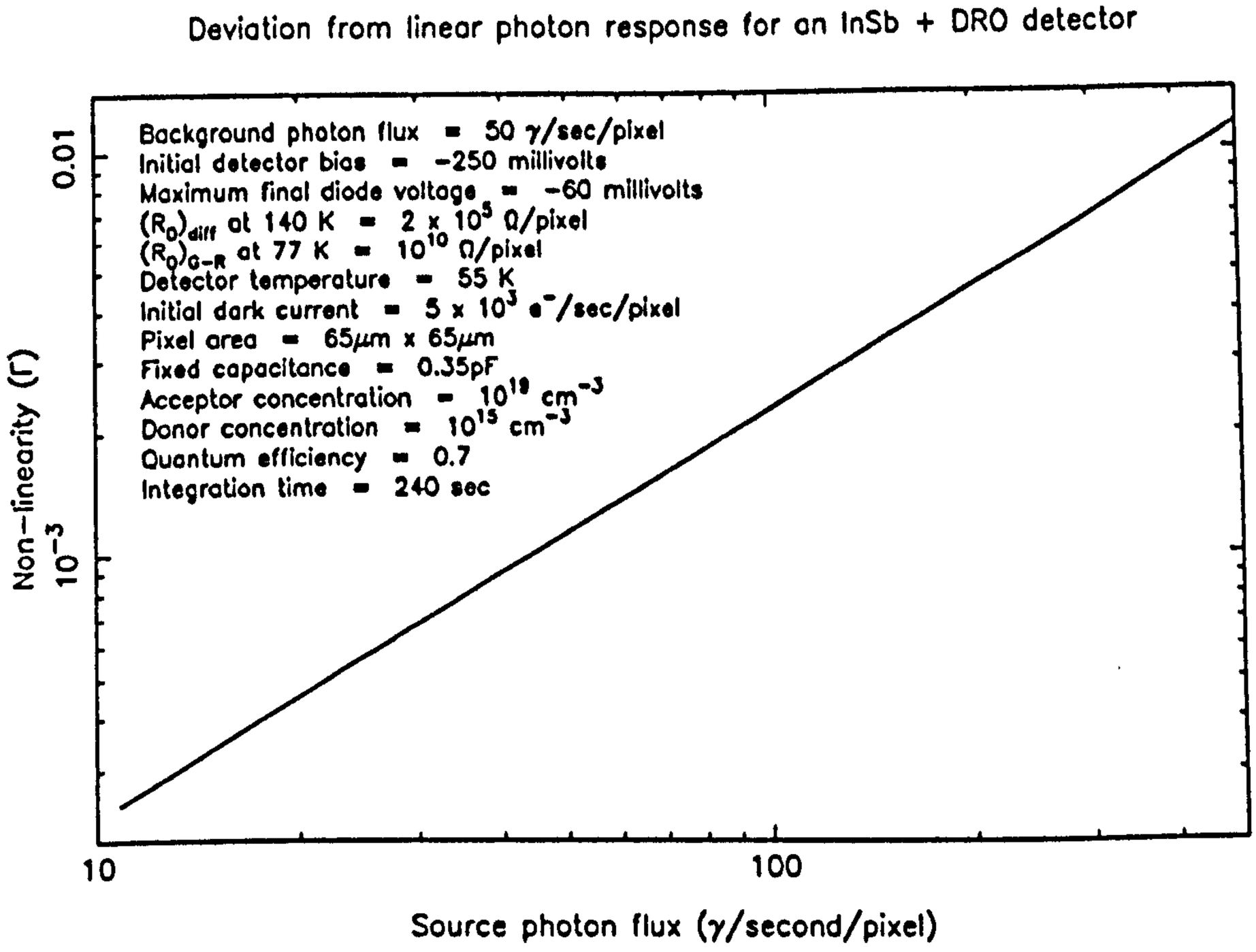


Figure 3.28: Non-linearity versus source flux — Regime 4

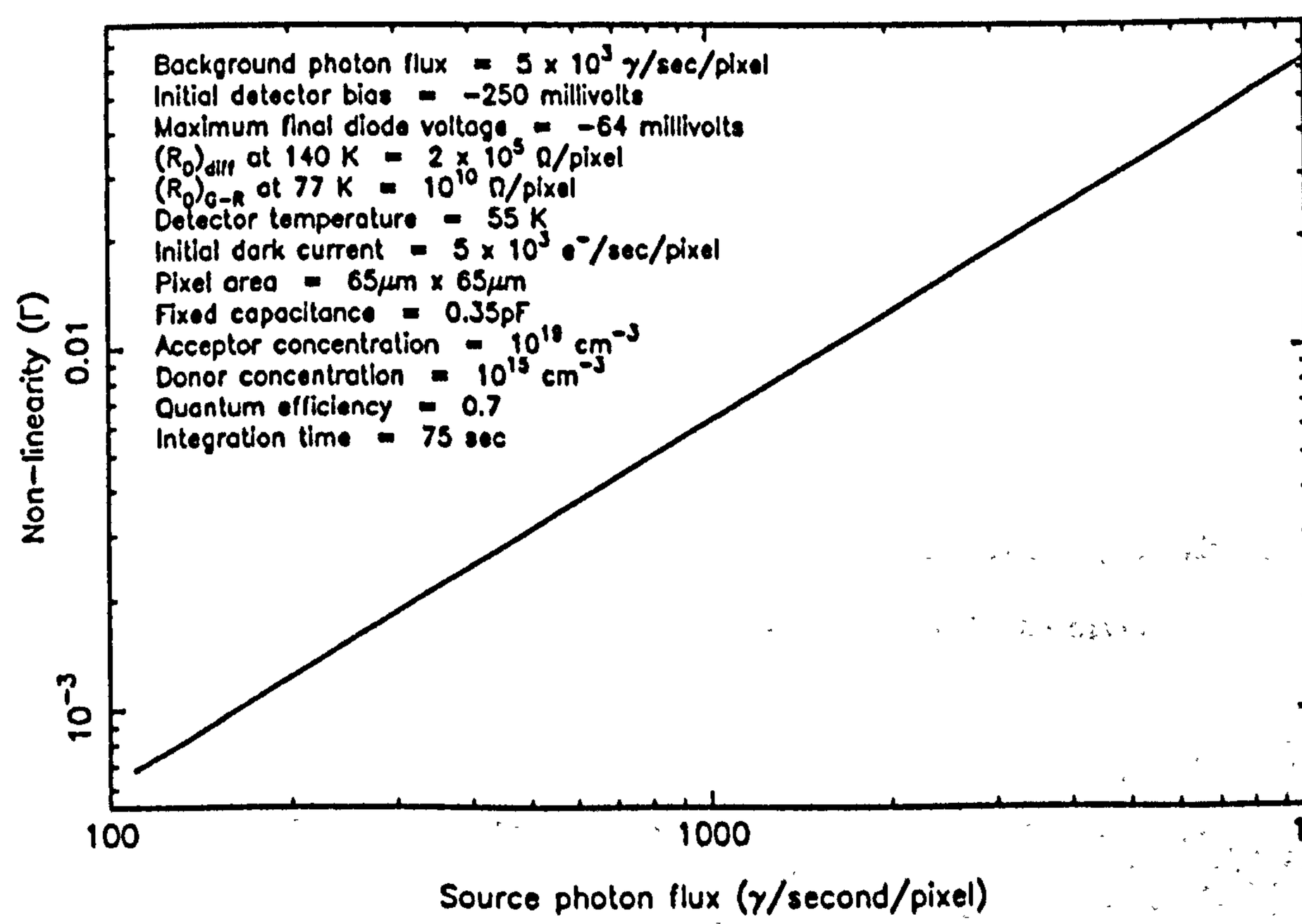


Figure 3.29: Non-linearity versus source flux — Regime 5

is the non-linearity, at up to 10%.

- In a mixed situation (Regime 5), non-linearities anywhere from 0.1 to 10% can be anticipated.

In the following section, we shall examine two algorithms that may be used to compensate for the non-linearity discussed here. We shall examine their effectiveness in removing the non-linearity seen in Regimes 1 to 5, and the advantages and disadvantages that arise when applying practical implementations of the correct algorithms to real data.

3.7 Linearity correction

In this section, we shall discuss two separate approaches to solving the non-linearity problem. Each approach has distinct advantages and disadvantages, and we shall attempt to evaluate algorithms corresponding to each approach, in an effort to determine the most suitable calibration procedure for practical use. Both approaches and general algorithms are due to Alan Hoffman of SBRC [66], but the specific algorithms, their implementations, and application to model and actual data are due to this author.

Firstly we shall derive the two solutions, and then we shall examine their effectiveness in removing the non-linearity seen in each of the five regimes discussed in the previous section. We shall also discuss the relative merits and dismerits of each solution, in the context of compensating for the non-linearity of real data.

A third approach to correcting for the non-linearity of the SBRC array has been discussed by Fowler *et al.* [42], but this technique corresponds to the analytical solution discussed later in this section, in the limiting case of zero dark current.

3.7.1 A general numerical solution

The complex equation derived to describe the voltage discharge of the unit cell proves very difficult to solve analytically for the photon flux ϕ, the desired quantity. However, we may be able to obtain a numerical solution as follows.

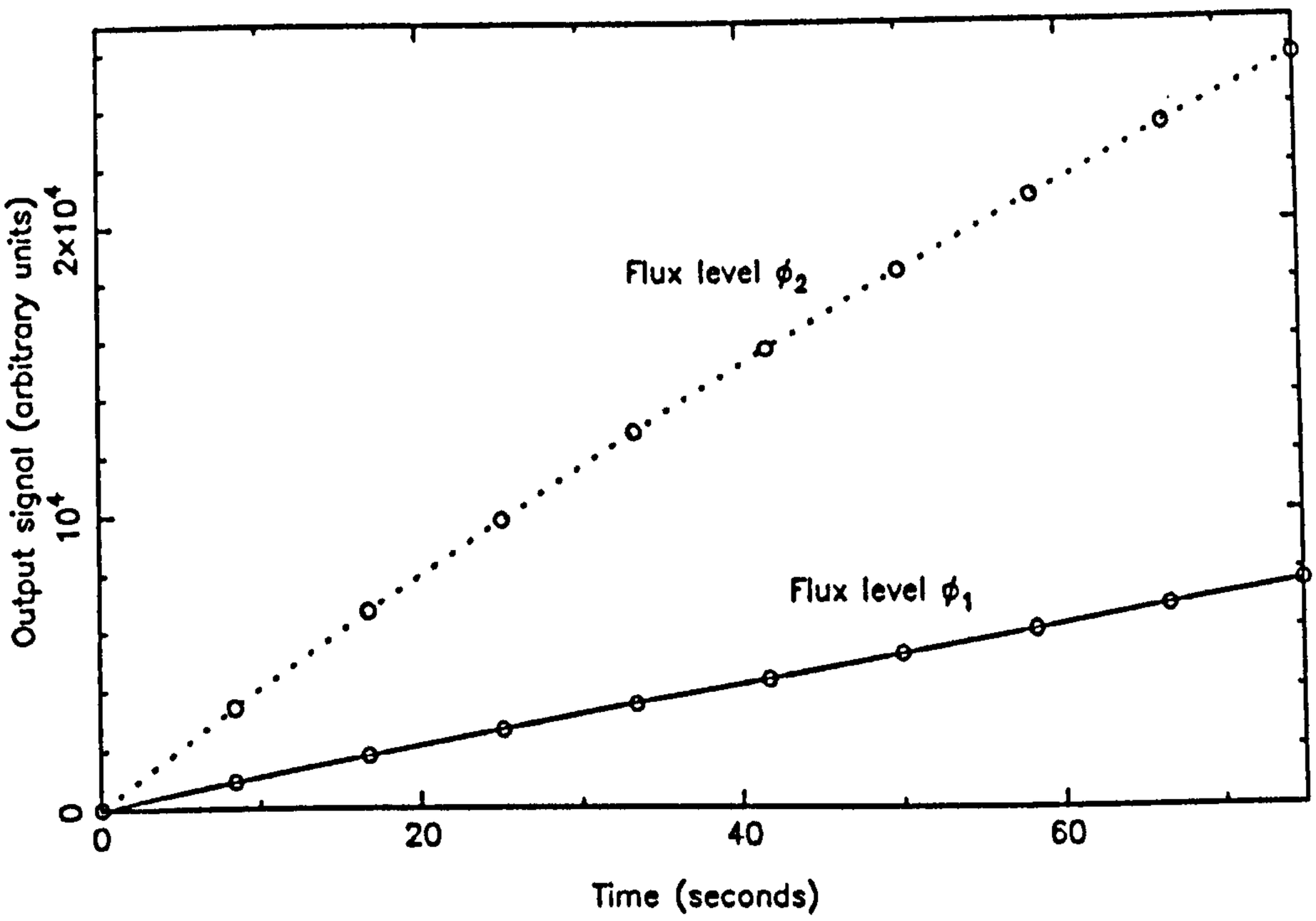


Figure 3.30: Measurement of S versus t for ϕ_1 and ϕ_2

We can represent the the voltage discharge rate dV/dt as a linear superposition of a voltage dependent dark current term $h(V)$ and a photo-current term, comprising of the photon flux ϕ and a voltage dependent photon gain term $g(V)$. This gives us the general representation

$$\frac{dV}{dt} = h(V) + g(V)\phi \tag{3.52}$$

In fact, as we will be deriving a general numerical solution, we can replace the specific notion of voltage V with a more generalised output signal S, which may be the number of ADUs stored by the computer. This gives us :

$$\frac{dS}{dt} = h(S) + g(S)\phi \tag{3.53}$$

We make no assumptions as to the theoretical forms of $h(S)$ and $g(S)$: that is, we make no *a priori* assumptions about the type of detector being used.

We calibrate as follows. Taking a fixed initial bias, and a fixed detector temperature (ensuring a fixed dark current term), we measure the output signal S as a function of time t, for two fixed illumination levels ϕ_1 and ϕ_2, as represented schematically in figure 3.30. Using a polynomial fitting algorithm, these data may be reduced to give

$$\left.\frac{dS}{dt}\right|_{\phi_1} \quad \text{and} \quad \left.\frac{dS}{dt}\right|_{\phi_2}$$

as functions of time t. However, what we want is these as functions of the output signal S, which shall assume to be possible at this stage, although we shall discuss this point in detail later. This allows us to write the two equations :

$$\left.\frac{dS(S)}{dt}\right|_{\phi_1} = h(S) + g(S)\phi_1 \tag{3.54}$$

$$\left.\frac{dS(S)}{dt}\right|_{\phi_2} = h(S) + g(S)\phi_2 \tag{3.55}$$

Manipulating these equations, we can derive the photon gain term $g(S)$ and the dark current term $h(S)$ as follows :

$$g(S) = \left(\left.\frac{dS(S)}{dt}\right|_{\phi_2} - \left.\frac{dS(S)}{dt}\right|_{\phi_1}\right)(\phi_2 - \phi_1)^{-1} \tag{3.56}$$

$$h(S) = \left(\left.\frac{dS(S)}{dt}\right|_{\phi_1}\phi_2 - \left.\frac{dS(S)}{dt}\right|_{\phi_2}\phi_1\right)(\phi_2 - \phi_1)^{-1} \tag{3.57}$$

Thus, we can now use these expressions for $g(S)$ and $h(S)$ to numerically integrate dS/dt for any given integration time and photon flux in the following fashion :

$$\begin{aligned} S(t=0) &= S_0 \\ S(\delta t) &= S_0 + [h(S_0) + g(S_0)\phi]\,\delta t \\ S(T+\delta t) &= S_T + [h(S_T) + g(S_T)\phi]\,\delta t \end{aligned}$$

In this way, a series of numerical integrations would be made, for a range of photon fluxes, each with the same integration time. These data would be fitted to obtain a function $\phi = \phi(S)$, such that the input photon flux can be derived from the output signal for the given integration time.

This technique has the apparent problem that two calibration flux levels (ϕ_1 and ϕ_2) must be used, and their (relative) brightnesses known. This would prove difficult to provide in practice, with any degree of accuracy. However, by setting ϕ_1 to zero by blanking the detector off and taking a series of dark current exposures, and by replacing ϕ_2 with some unknown flat-field flux ϕ_f, the forms of $g(S)$ and $h(S)$ reduce to :

$$g(S) = \left(\left.\frac{dS(S)}{dt}\right|_{\text{dark}} - \left.\frac{dS(S)}{dt}\right|_{\phi_f}\right)\frac{1}{\phi_f} \tag{3.58}$$

$$h(S) = \left.\frac{dS(S)}{dt}\right|_{\text{dark}} \tag{3.59}$$

Thus, after the numerical integration procedure described above, in which the output signal S is calculated for a range of input photon fluxes ϕ at one fixed integration time,

these fluxes are known only as ratios of the unknown flat-field flux (ϕ/ϕ_f). Only a relative calibration is obtained. However, this is the common astronomical situation, and is acceptable. Absolute calibration takes place by similar observations of standard stars.

It is important to note that many photon fluxes can be calibrated from several measurements of one single fixed flux via the assumption that both $h(S)$ and $g(S)$ depend on the signal level only, and not on the actual value of the photon flux. That is, the two functions depend only on how much of the well has been filled, and not how rapidly it got to that state.

3.7.1.1 Practical implementations

The advantage of this numerical technique is that $\hat{it}$ is completely generalised. No *a priori* knowledge about the forms of $h(S)$ and $g(S)$ is assumed from knowledge of the detector physics. Therefore, this method may be applied to any detector type. The accuracy of the technique is discussed in section 3.7.3 below.

There are however two major disadvantages. Firstly, a lot of data must be collected in order to ensure accurate determinations of the $h(S)$ and $g(S)$ functions : images at many different integration times for both the dark current and flat-field cases must be obtained. This may prove time consuming, although for a stable detector system, the calibration frames could be obtained at dusk and dawn. Secondly, the reduction of the data to the $h(S)$ and $g(S)$ functions, the subsequent numerical integrations of $dS(S)/dt$ for a range of photon fluxes, and the final fit of $\phi = \phi(S)$, all take considerable amounts of computer time, as the whole procedure must be carried out for each pixel in the detector array individually. Timing tests are discussed below.

In a practical implementation of this algorithm, one way of minimising the necessary computer time is to use polynomial representations of the various functions throughout. This proves difficult almost at the first hurdle. As discussed above, the first step is to obtain two series of timed exposures, one with no illumination, and the other illuminated by a flat-field. In each case, the data may be fitted with a polynomial, to give the function $S = S(t)$. However, at the next stage, it is not $dS(t)/dt$ that is required, but $dS(S)/dt$ (see equations 3.54 and 3.55). This is because we wish to know the rate

of change of signal as a function of the signal itself, not as a function of the integration time. To obtain $dS(S)/dt$ from $S(t)$ for a general polynomial fit is not simple — $dS(t)/dt$ is evaluated, and for a number of values of t, both $S(t)$ and $dS(t)/dt$ are numerically evaluated, and then a fit is made to the resulting data, giving a polynomial fit to $dS(S)/dt$. This can be inaccurate if not enough values of t are chosen, but time consuming if many are chosen.

An alternative technique can be used that involves manipulation of polynomial coefficients only. In this way, the derivation of $dS(S)/dt$ for the dark and flat-field cases, and thence $h(S)$ and $g(S)$ is simpler and faster. However, it does involve the assumption that the initial polynomial fit is sufficiently accurate at second-order only. Firstly, the signal S is fitted as a function of the time t :

$$S(t) = a + bt + ct^2$$

This is differentiated with respect to t to give :

$$\frac{dS(t)}{dt} = b + 2ct$$

Then the same data are fitted the other way, to give the integration time t as a function of the signal S :

$$t(S) = a' + b'S + c'S^2$$

Substituting for t gives us :

$$\begin{aligned} \frac{dS(S)}{dt} &= b + 2c(a' + b'S + c'S^2) \\ &= (b + 2ca') + 2cb'S + 2cc'S^2 \\ &= A + BS + CS^2 \end{aligned}$$

where the final polynomial coefficients A, B, C are evaluated merely by manipulating the original coefficients of $S(t)$ and $t(S)$. These derived coefficients are then manipulated to form the polynomial coefficients of the two functions $h(S)$ and $g(S)$, according to equations 3.56 and 3.57.

Whilst this small trick speeds up the overall initial reduction process (from raw data to $h(S)$ and $g(S)$ functions), the major time penalty is still incurred in the second procedure, the numerical integration of S from $h(S)$ and $g(S)$ for a number of photon fluxes.

3.7.1.2 Timing tests

As can be seen from the discussion above, any practical linearisation procedure based on this general approach logically divides into three routines. The first routine would derive the $h(S)$ and $g(S)$ functions from the raw data. One such derivation per night's data should be enough. The second routine determines the function $\phi = \phi(S)$ for a given integration time, by generating a series of numerical integrations with a range of input photon fluxes for that particular integration time. These data would then be fitted with a polynomial to give the required curve. This procedure would be lengthy, but would necessarily be repeated each time the on-chip integration time was changed, potentially up to fifty times per night's data. The final process of deriving the input flux from the output signal will be quite fast, as it involves a trivial manipulation of the polynomial coefficients of the applicable $\phi(S)$ function, and would be necessary for each individual frame read out of the array.

In order to assess the speed of this generalised algorithm, three routines were coded to perform the three steps discussed above (GENCAL1, GENCAL2, and GENCAL3). Each algorithm works on a per pixel basis, and therefore its speed depends on the size of image being handled. Throughout, we assume an image size of 62 × 58 pixels.

Using a routine (NL) based on the non-linear discharge equation derived earlier in this chapter (equation 3.43), the output signal as a function of time was simulated for 2 sets of 10 calibration images, one integrating on dark current alone, and the other on a flat field source. These images were input to GENCAL1, which then reduced them to the polynomial coefficients of the $g(S)$ and $h(S)$ functions. Using a second order polynomial, this fitting procedure took 51 seconds of Vax 11/780 CPU time.

The routine GENCAL2 was then used to numerically integrate the $g(S)$ and $h(S)$ functions for a range of input fluxes, each at the same fixed integration time. The speed of this process depends heavily on the number of fluxes considered, and the number of steps used in the numerical integration. From tests, 10 flux levels and 25 integration steps are found to yield sufficiently accurate results for a well-behaved non-linearity. Using these numbers, GENCAL2 took 475 seconds of 11/780 CPU time to calculate the output signals at each flux level, and an additional 11 seconds of CPU time to fit a second order polynomial to the each pixel, yielding the relative photon flux as a function

of output signal, for the given fixed integration time.

Finally, GENCAL3 took these polynomial coefficients as input, and an image with the same fixed integration time, again simulated by NL. The routine converted the image from output signal units to relative photon units in less than 0.5 seconds of CPU time.

To summarise :

- the initial calibration data could be obtained once per night if the detector system remained stable and at a constant temperature — the reduction to the $g(S)$ and $h(S)$ functions would be performed once only, and would take $\sim$ 1 minute of CPU time.

- numerically integrating the $g(S)$ and $h(S)$ functions against photon flux for a given integration time takes $\sim$ 10 minutes of CPU time, and would have to be performed once for each time the on-chip integration time was changed during a night.

- the final conversion of each image from output signal units to relative photon units is fast, and adds no overheads when compared to more traditional reduction procedures.

3.7.2 A specific analytical solution

Although the numerical solution discussed in the previous section has the benefit of making no *a priori* assumptions about the physics that determines the detector discharge, the method requires both a large amount of calibration data and a large amount of computer time in order to apply the linearity correction. As mentioned before, the general detector discharge equation is not directly solvable, but some specific forms of the equation do exist for which it is possible to derive an analytical solution. Such equations, by necessity, make *a priori* assumptions about the nature of the physics governing the detector, thus reducing the effective application of any solution to detectors which obey these assumptions. However, we shall see that any loss of generality is more than compensated for by the speed with which such an algorithm will calibrate an image.

We shall now examine a particular solution applicable to the SBRC InSb + DRO arrays; that is, photo-voltaic detectors that self-integrate. Again, the basic algorithm is due to Alan Hoffman of SBRC [66], but the implementation and analysis are due to this author. It should be noted that Hoffman's paper on this topic contains several typographical errors — we believe the discussion presented here is correct.

We begin by rewriting equation 3.53 in terms of the rate of change of the charge on a detector :

$$-\frac{dQ}{dt} = I_{dark} - I_{photon} \tag{3.60}$$

where Q is the charge on the detector capacitance C, and the photo-current $I_{photon} = \eta q \phi$. Note the sign convention : In reverse bias, Q and I_{dark} are negative, and both the dark and photo-current terms act to discharge the detector, i.e. $Q \rightarrow 0$.

The form of the capacitance is assumed to be as given in equations 3.36 and 3.38, in terms of its fixed and bias dependent components, as :

$$C = C_{fix} + C_0\left(1 - \frac{V}{V_{bi}}\right)^{-\frac{1}{2}} \tag{3.61}$$

where C_{fix} is the fixed (bias independent) part

C_0 is the zero bias junction capacitance

As we have seen, it is the change in the junction capacitance as a function of the reverse bias that is the main source of the detector non-linearity, and a major assumption of this analytical calibration technique is that we can determine its form by measurement or analysis. We shall discuss this further below, but at this point, we assume we know the form of $C = C(V)$.

Before proceeding, we make the following definitions. We define three quantities S_S, S_F, and S_D, as the signal (in data numbers for example) measured at the output of the camera system after equal integration time exposures on the source, a flat field, and blanked off, respectively. Assuming a known initial bias of V_i, a total gain G in DN/volt (i.e. the detector MOSFET gain times the system electronic gain), we can calculate the corresponding final detector voltages (V_S, V_F, V_D) for each exposure from :

$$\begin{aligned} S_S &= G(V_S - V_i) \\ S_F &= G(V_F - V_i) \end{aligned} \tag{3.62}$$

$$S_D = G(V_D - V_i)$$

noting that V_i, V_S, V_F, V_D are all negative. These may be rearranged to give :

$$\begin{aligned} V_S &= (S_S/G) + V_i \\ V_F &= (S_F/G) + V_i \\ V_D &= (S_D/G) + V_i \end{aligned} \tag{3.63}$$

We now define the charge stored on the detector capacitance at the beginning of each integration as Q_i, from the initial bias and our assumed knowledge of the form of the detector capacitance :

$$Q_i = V_i \times C(V_i)$$

and we similarly define the charge remaining on the detector after each integration as :

$$\begin{aligned} Q_S &= V_S \times C(V_S) \\ Q_F &= V_F \times C(V_F) \\ Q_D &= V_D \times C(V_D) \end{aligned} \tag{3.64}$$

With these definitions, we now discuss two specific cases; one with a constant, bias independent dark current, and one with a generation-recombination (G-R) type dark current.

3.7.2.1 Bias independent dark current case

Here we shall assume that the dark current is constant throughout the discharge. This may be true if the dark current is dominated by surface effects, or if it is actually a low level residual photon leak. For constant dark and photo-current, the charge remaining after an integration time T is given by :

$$\begin{aligned} Q &= Q_i + T\frac{dQ}{dt} \\ &= Q_i - T(I_{dark} - I_{photon}) \\ &= Q_i - T(I_{dark} - \eta q \phi) \end{aligned}$$

Thus, if the source and flat fields have photon fluxes ϕ_S and ϕ_F respectively, and for the dark current exposure $\phi_D = 0$, then we have :

$$Q_S = Q_i - T(I_{dark} - \eta q \phi_S)$$

$$Q_F = Q_i - T(I_{dark} - \eta q \phi_F) \tag{3.65}$$
$$Q_D = Q_i - T I_{dark}$$

This equations can be trivially manipulated to obtain the linear ratio of the source flux to the flat field flux as :

$$\frac{\phi_S}{\phi_F} = \left(\frac{Q_S - Q_D}{Q_F - Q_D}\right) \tag{3.66}$$

In practice, we would probably rescale the resulting ratio by the median of the level in the dark current subtracted flat field frame, in order that source frames calibrated with different level flat fields could be more easily compared. Before discussing this result any further, we shall proceed with the more complex case of a G-R type dark current.

3.7.2.2 Generation-recombination dark current case

The constant dark current case examined above may apply in very low temperatures regimes, but between 40 and 120 K, it is expected that the generation-recombination (G-R) dark current mechanism will dominate diffusion, tunneling, and surface effects. Recalling equation 3.18, the form of G-R current is given by :

$$I_{G-R} = \frac{2V_t}{R_0}\left(1 - \frac{V}{V_{bi}}\right)^{\frac{1}{2}}\left(e^{\frac{V}{2V_t}} - 1\right) \tag{3.67}$$

where $R_{0_{G-R}}$ is the applicable value of the zero bias dynamic impedance, R_0. For a reverse bias of ≥ 40 mV, the term $e^{V/2V_t} \rightarrow 0$, and by expanding the term $(1 - V/V_{bi})^{1/2}$, we obtain an approximation to the G-R current as :

$$I_{G-R} = -\frac{2V_t}{R_0}\left(1 - \frac{V}{2V_{bi}}\right) \tag{3.68}$$

Now, in order to provide a solution to equation 3.60, we need a term in the charge Q in the expression for the dark current, rather than in the voltage V. In order to do this, we make a substitution into the G-R current equation that assumes $C(V) = C_0$ for all V, where $C_0 = C(V = 0)$. This appears somewhat strange, as it is the very fact that $C(V) \neq C_0$ that is the source of the non-linearity. In fact, it is simply an *ad hoc* assumption which leads to an easily solved differential equation, and no physical significance is attached (Hoffman, private communication). However, Hoffman has shown that making this assumption leads to an error in the resulting G-R current compared to the true G-R current of only $\sim 5\%$ between 25 and 275 mV reverse bias [66].

Therefore, we proceed, changing equation 3.68 thus :

$$I_{G-R} = -\frac{2V_t}{R_0}\left(1 - \frac{C(V)}{C_0}\frac{V}{2V_{bi}}\right) \tag{3.69}$$

Then, recalling that $Q = C(V) \times V$, we can can rewrite this as :

$$I_{G-R} = -\frac{2V_t}{R_0}\left(1 - \frac{Q}{2C_0V_{bi}}\right) \tag{3.70}$$

Inserting this into equation 3.60, we now have a differential equation in the charge Q :

$$-\frac{dQ}{dt} = -\frac{2V_t}{R_0}\left(1 - \frac{Q}{2C_0V_{bi}}\right) - \eta q \phi \tag{3.71}$$

This differential equation may be solved (by using $\int (ax+b)^{-1}dx = \frac{1}{a}\log(ax+b)$) to give the charge Q_T remaining on the detector after an integration time T :

$$Q_T = \left(Q_i - C_0V_{bi}\left(\frac{\eta q \phi R_0}{V_t} + 2\right)\right)e^{-\frac{V_tT}{C_0V_{bi}R_0}} + C_0V_{bi}\left(\frac{\eta q \phi R_0}{V_t} + 2\right) \tag{3.72}$$

The actual details of this equation are not important. If we continue, by inserting $Q_T = Q_D$ for $\phi = 0$, and $Q_T = Q_S$ and $Q_T = Q_F$ for $\phi = \phi_S$ and ϕ_F respectively, we obtain the following :

$$Q_D = (Q_i - 2C_0V_{bi})\,e^{-\frac{V_tT}{C_0V_{bi}R_0}} + 2C_0V_{bi} \tag{3.73}$$

$$Q_S = \left(Q_i - C_0V_{bi}\left(\frac{\eta q \phi_S R_0}{V_t} + 2\right)\right)e^{-\frac{V_tT}{C_0V_{bi}R_0}} + C_0V_{bi}\left(\frac{\eta q \phi_S R_0}{V_t} + 2\right) \tag{3.74}$$

$$Q_F = \left(Q_i - C_0V_{bi}\left(\frac{\eta q \phi_F R_0}{V_t} + 2\right)\right)e^{-\frac{V_tT}{C_0V_{bi}R_0}} + C_0V_{bi}\left(\frac{\eta q \phi_F R_0}{V_t} + 2\right) \tag{3.75}$$

$$\tag{3.76}$$

These equations can be manipulated to give us the simple result that :

$$\frac{\phi_S}{\phi_F} = \left(\frac{Q_S - Q_D}{Q_F - Q_D}\right) \tag{3.77}$$

Thus, the same trivial calibration algorithm applied to the constant dark current case (equation 3.66) may also be applied to data in which generation-recombination is the dominant dark current mechanism. Therefore, the simple normalisation process discussed provides a fast method of linearising data from the SBRC array, under a range of likely circumstances. In the intermediate temperature range 40–120 K, G-R current is accounted for. Below 40 K, as long as the dark current is zero or constant with bias, the algorithm should also work. Even if the low temperature dark current is a residual photon leak, the same process will apply. Only if the dark current exhibits dependence on the bias other than that of G-R current (e.g. tunneling), will the algorithm be less

satisfactory. Note however that any such error will probably be small for ground-based imaging cameras at least, as the photon background should virtually always dominate the low temperature dark current, and because the algorithm will correct the fundamental non-linearity (i.e. the dependence of the capacitance on reverse bias) regardless of the form of the dark current.

3.7.2.3 Practical considerations

We can see that the specific algorithm outlined above is entirely analogous in form to the simple linear reduction procedure discussed in section 3.6.1, with the simple exception that we apply a bias dependent normalisation factor to each term, effectively linearising the data before we dark subtract and flat field. How is the simple algorithm implemented in practice? We require three basic data frames; the source frame, a flat field frame, and a dark current exposure. This must all start at the same initial bias, and have the same on-chip integration time, with the detector at a fixed temperature throughout. The latter two calibration frames ideally should be comprised of many more co-adds than the source frame, to reduce their noise contribution. This is probably feasible for the dark current frame, which can be obtained 'off-line' during dusk or dawn, assuming the detector remains stable. As seen elsewhere in this thesis however, the best flat fielding is obtained when using sky frames taken very close in time to the source frames, and once the number of sky and source co-adds are equal, additional noise reduction is hard to come by. At this point, equal numbers of sky and source co-adds are recommended, noting that in future, suitable median stacking algorithms may be used to generate a low noise 'master' sky flat from a whole nights data.

The reduction steps are as follows : Firstly, all three frames are divided by the relevant number of co-adds, to normalise them to one co-add. Then we must convert our output signals (in DN for example) into the final charge on the detector (Q_S, Q_F, Q_D) according to equations 3.62–3.64. In order to do so, we need to know V_i, the initial bias; G, the total system gain; and the form of the capacitance $C(V)$. The initial bias V_i should be known, but losses in the external electronics and the detector reset may result in uncertainties of $\sim 10\,\%$. The gain G is comprised of two components; the mean detector MOSFET gain, and the electronic gain in converting the array output voltage into data numbers. Both should be relatively easy to measure accurately. Finally, the

form of $C(V)$ must be known. We can determine it either by measurement or analysis. In the former case, we can take a series of exposures on a fixed brightness source, with a range of integration times. Converting to voltage via V_i and G as above, we can obtain a polynomial fit to the rate of change of voltage as a function of the voltage. Provided the experiment is heavily photon dominated, this curve will effectively give us $C(V)$, as $dV/dt = K/C(V)$, where K is the constant photo-current. Alternatively, we can calculate the form of $C(V)$ analytically, by assuming it comprises a known fixed part C_{fix}, and a bias dependent junction capacitance C_{jn} which we can calculate from equation 3.38, provided we know the detector area, temperature, and doping levels. After converting all our output signals into charges, we dark subtract and flat field as in equation 3.66, probably followed by a rescaling by the median dark subtracted flat field value, as mentioned previously. The result should be an image with the value of each pixel directly proportional to the photon flux falling on it.

Whilst it would appear that the number of *a priori* assumptions that must be made about the specific detector parameters would tend to make the final linearisation process prone to error, it can be seen that these parameters tend towards first order cancellation in equation 3.66. Hoffman [66] has shown that this is indeed the case. In a simple error analysis, he shows that for a 10% error in the assumed value of V_i for example, an change in the output flux ratios of less than 1% would be expected for a wide range of fluxes. The accuracy of the technique is assessed in section 3.7.3 below.

3.7.2.4 Timing tests

In order to assess the speed of this technique, a routine (SBRC_LIN) was implemented. This algorithm takes as input raw source, flat field, and dark frames, along with values of the initial bias, MOSFET gain, and detector temperature. From the estimated bias and gain, the algorithm converts each frame to on-chip voltage units. Then, from the temperature and the SBRC array modelling software used throughout this chapter, the algorithm calculates the theoretical capacitance as a function of voltage, in order to convert from voltage to charge units. Then the simple algorithm given in equation 3.66 is applied, followed by a rescaling. The output result is an image calibrated in relative photon units.

Apart from the conversion from raw signal to charge, this procedure is identical to the usual linear reduction case, and therefore takes only a little longer to execute. The routine SBRC_LIN was timed to take 1 second of Vax 11/780 CPU time to calibrate a given 62 × 58 pixel image. As the same dark and flat field frames are likely to be used to calibrate several different source frames, some savings could be made by pre-processing the frames into charge units in a separate routine. This saving is minor, and not considered worthwhile for interactive use.

3.7.3 Accuracy of the linearisation algorithms

In this section, we shall briefly examine the accuracy of the two algorithms when applied to data from the five different observing regimes discussed in section 3.6.2.

For each of the five regimes, we have taken the same parameters used to derive figures 3.25 to 3.29 and have then used the two algorithms to linearise them. The implementations of the algorithms are identical to those discussed above, in sections 3.7.1 and 3.7.2, but working on a single pixel basis only. It should be noted that in the case of the analytical algorithm, the required values of the initial reverse bias, gain, and dependence of the capacitance on reverse bias, were all made equal to the actual values used by the model. Thus the linearity correction seen here represents the ideal case, and actual results may be somewhat worse.

Finally, non-linearity (Γ) is derived *before* correction, and then *after* correction by each algorithm. The results of these simulations are presented in figures 3.31 to 3.35.

3.7.3.1 Results

The first point arising from the figures is that both algorithms do indeed reduce the non-linearity from its uncorrected value. The second point however, is that neither algorithm is consistently better than the other.

In Regime 1, the thermal imaging case, the source fluxes are only minor perturbations on the background flux, and the uncorrected non-linearity is about 0.2% at most. The analytical technique slightly reduces the non-linearity, but the numerical technique

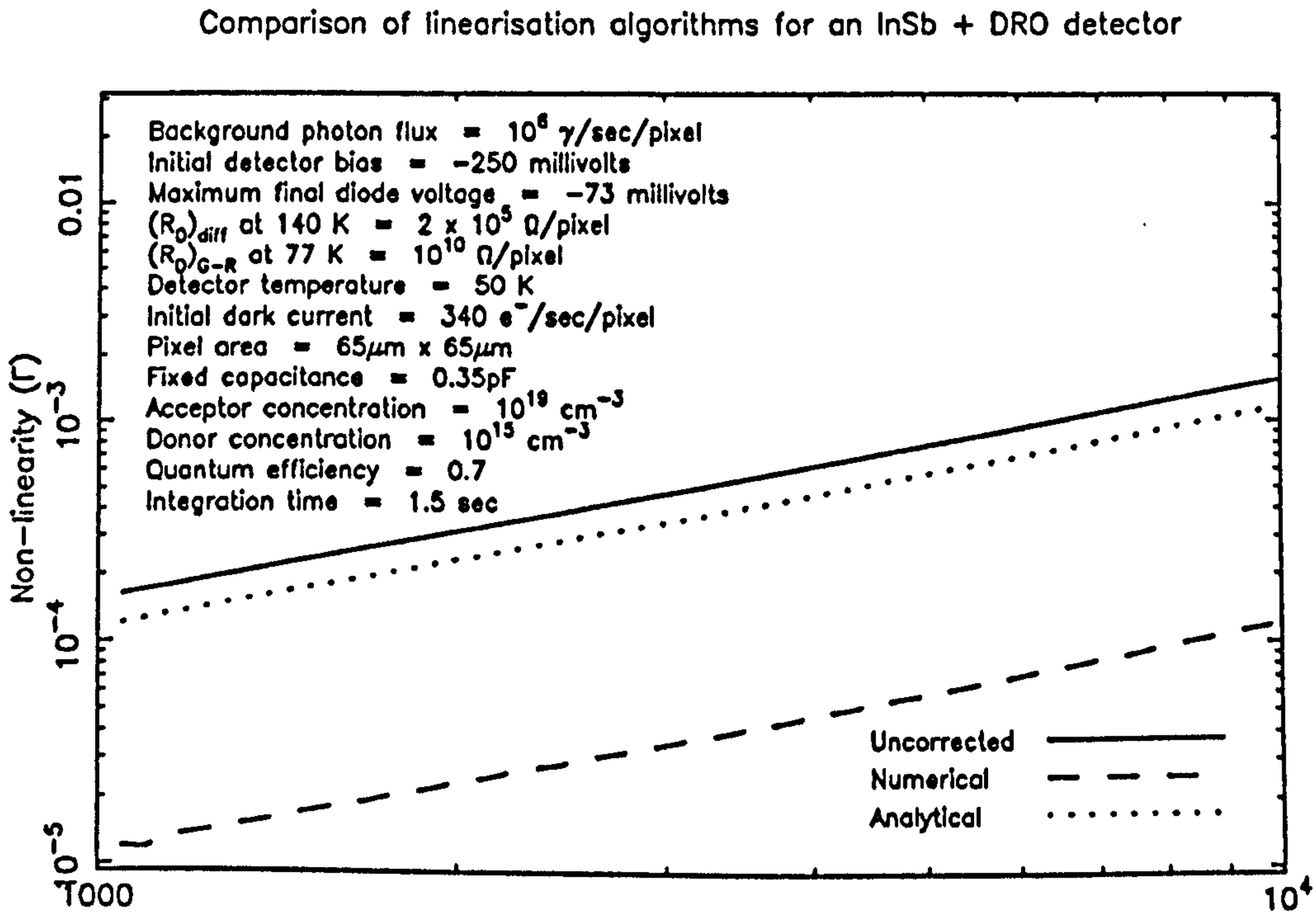


Figure 3.31: Comparison of linearisation algorithms — Regime 1

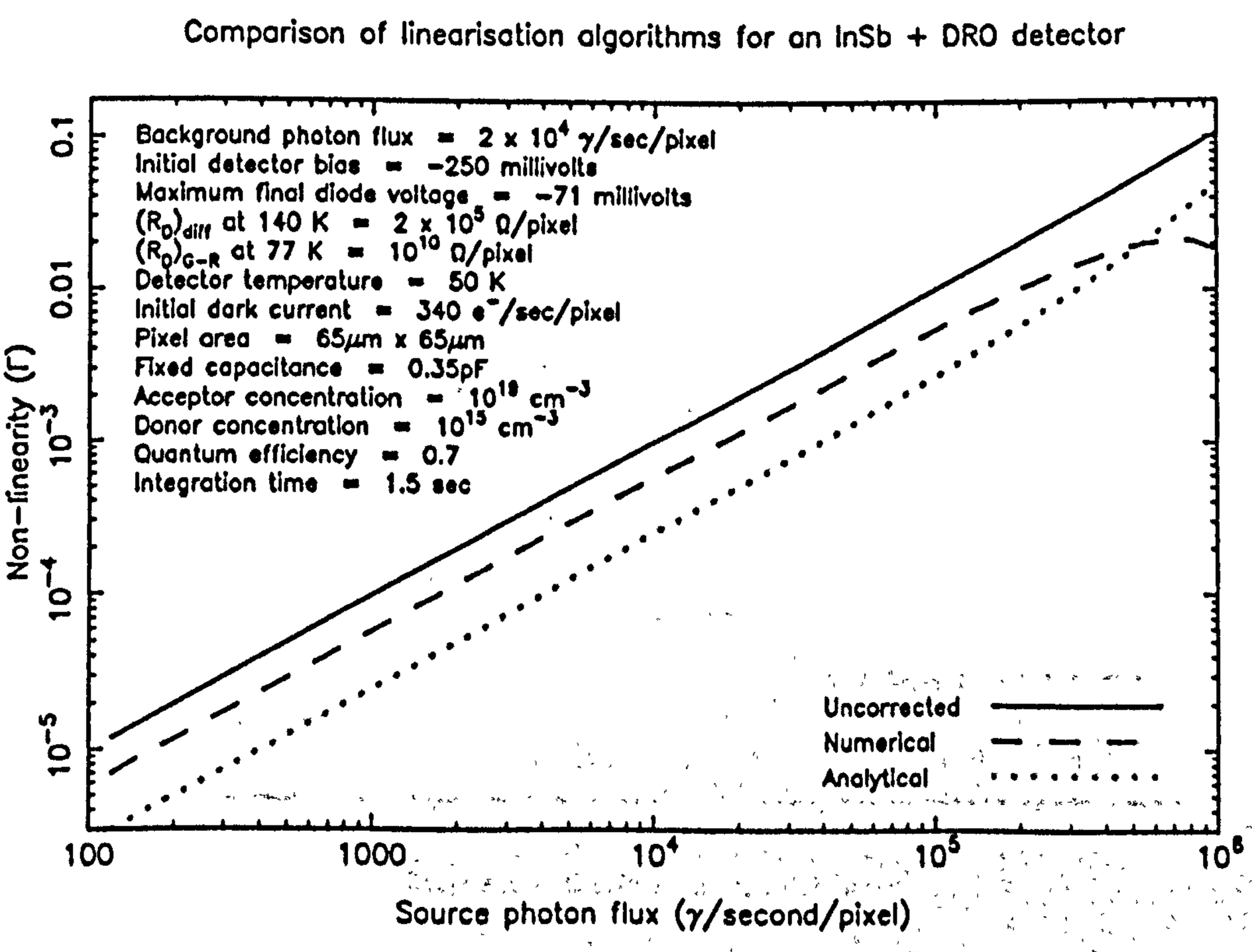


Figure 3.32: Comparison of linearisation algorithms — Regime 2

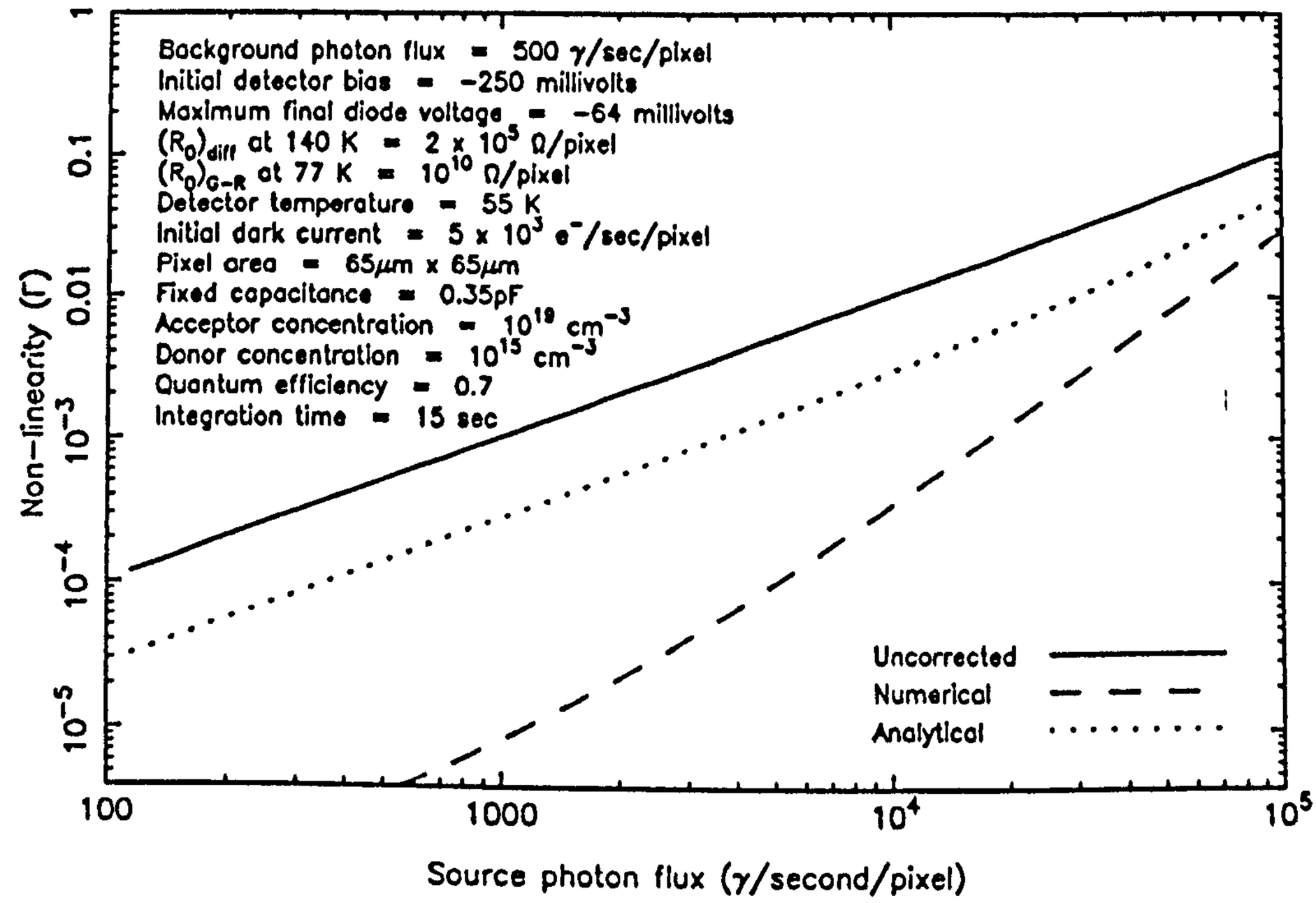


Figure 3.33: Comparison of linearisation algorithms — Regime 3

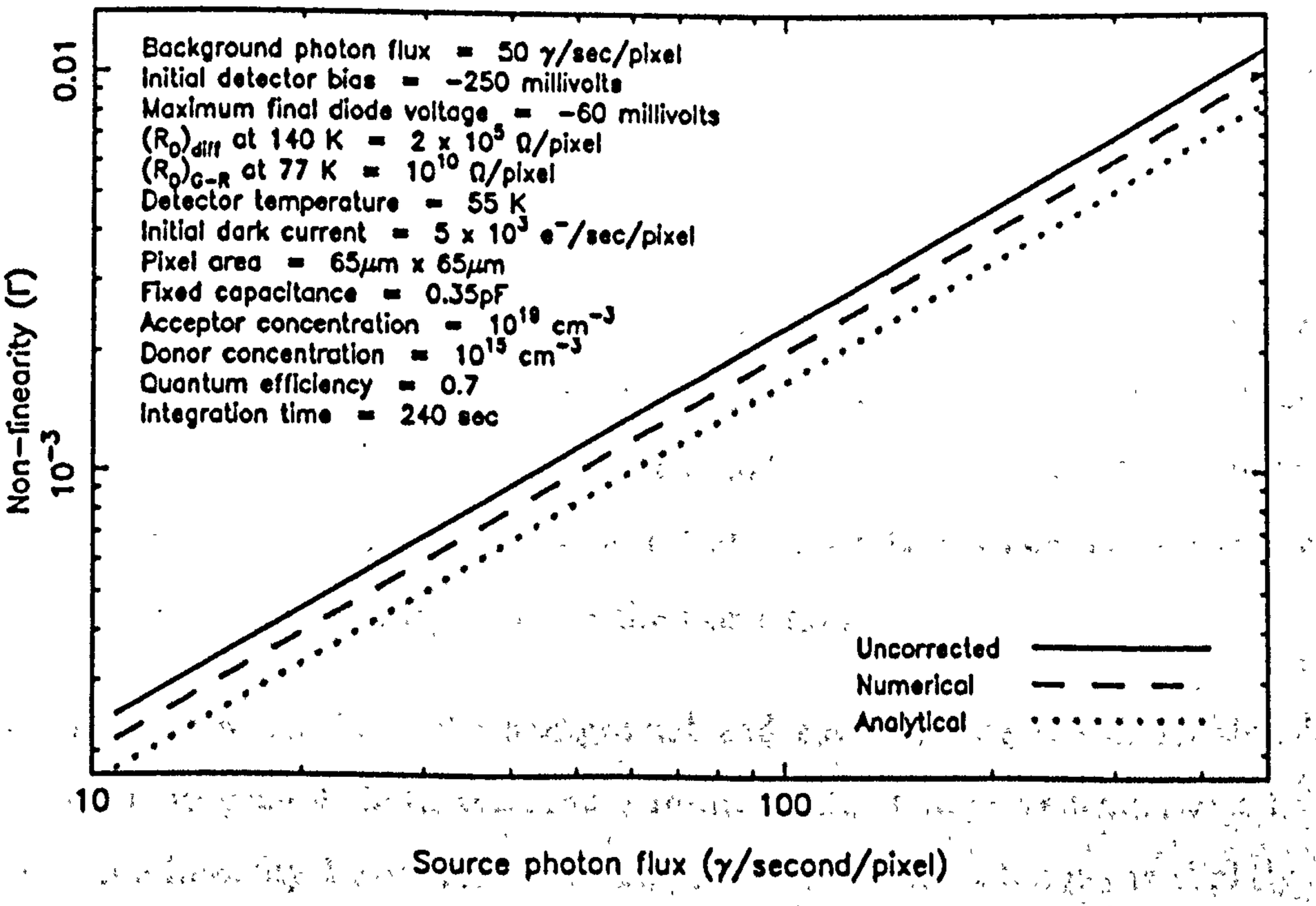


Figure 3.34: Comparison of linearisation algorithms — Regime 4

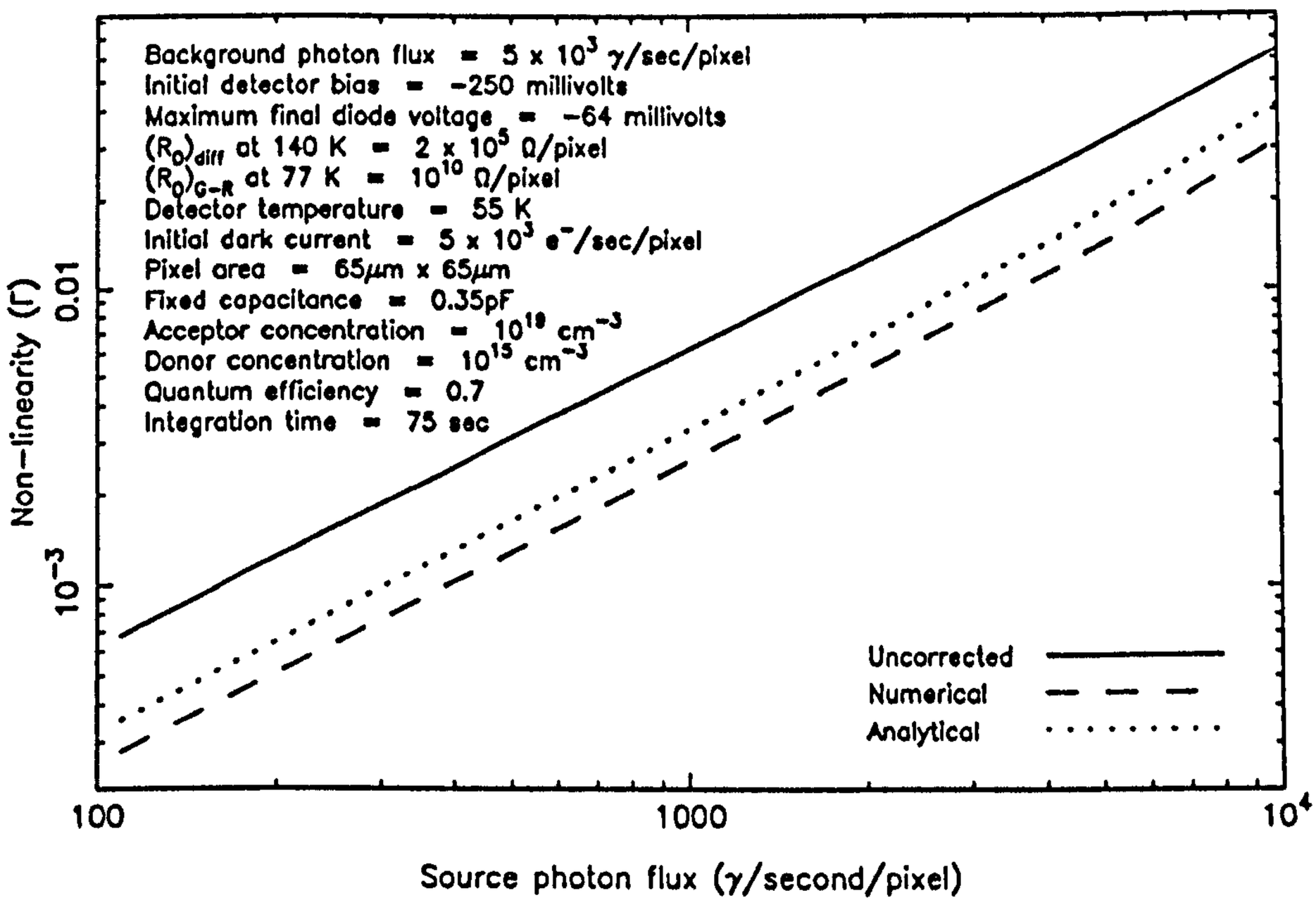


Figure 3.35: Comparison of linearisation algorithms — Regime 5

reduces it by an order of magnitude.

In the non-thermal broad band imaging case Regime 2, the source fluxes dominate the dark current and background. Although the analytical case is better over much of the source flux range, it is important to note that when the uncorrected non-linearity exceeds 10%, the analytical algorithm leaves a residual non-linearity still as large as 5%, whilst the numerical algorithm is never worse than 2%.

Regime 3 is the high spectral resolution imaging case, with low background, a moderate dark current, and a large range of source fluxes. The uncorrected non-linearity again exceeds 10% at worst, and again, the analytical solution can do no better than a 5% residual non-linearity. The numerical algorithm is consistently better, but still leaves a residual non-linearity of 3% in the worst case.

In the case where both the background and source fluxes are dominated by the background, Regime 4, both techniques result in slight improvements, and keep the residual non-linearity below 1% at worst. The analytical technique is slightly better than the numerical.

Finally, in Regime 5, the mixed case, where all three sources of current are roughly

equal, the numerical technique is better, keeping the residual non-linearity below 3% when the uncorrected value has reached 6%. The analytical technique is only slightly worse, resulting in a maximum residual non-linearity of 4%.

3.7.3.2 Discussion

Neither of the two techniques is far better than the other. The numerical technique is the most accurate of the two in the situations where the uncorrected non-linearity is worst, and manages to keep the maximum residual non-linearity at or below 3% in all cases. The analytical technique results in a maximum error of 5% in the broad band imaging regime (Regime 2), but gives fairly similar results to the numerical technique elsewhere.

The reasons for the somewhat inconsistent results are unclear. Certainly the assumptions made by the analytical algorithm, and the low degree polynomial fit used by the numerical algorithm means that neither algorithm can be expected to provide perfect calibration. Of the two techniques, the numerical algorithm has the most potential for improvements in accuracy, by taking more calibration data, and using a higher degree polynomial fit. This would further increase the already considerable reduction time however.

3.7.4 Summary of results from linearisation tests

In this section, we have derived a realistic non-linearity parameter, that describes the non-linearity *after* full data reduction. The SBRC array is seen to result in values of 10% in the worst cases, and generally less than 1%. We have also discussed two separate approaches to non-linearity correction. Both approaches have been implemented and tested on realistically modelled data from the SBRC array.

The numerical technique has the advantages of being completely detector independent, free of *a priori* assumptions, and able to keep residual non-linearity at or below 3% across the wide range of observational regimes. For many situations, the residual non-linearity is less than 1%. This approach has the disadvantage of requiring a fairly large amount of calibration data, and the use of a large amount of computer time in

the several steps taken to linearise a given image. Nevertheless, the latter disadvantage is not fundamental, and a better optimised implementation and faster computer would make this the preferred technique, due to its generality and potentially higher accuracy.

The analytical technique is much faster than the numerical technique, and in the ideal case, not much less accurate. It too keeps the residual non-linearity to less than 1% in many situations, although a maximum non-linearity of 5% is seen in some extreme cases. However, the technique has the disadvantage of making *a priori* assumptions about the physics of the detector involved, and is therefore only really applicable to reverse biased photovoltaic detectors. It also requires that a user make estimates of several important parameters, and even though errors in these parameters tend to cancel, any such errors will still degrade the results presented here, which were derived for the ideal case.

To summarise, we can see that the inherent non-linearity of the SBRC array may result in errors of up to 10% in the relative brightness of two sources within the same image. However, this maximum value occurs only in the extreme situation where one source is almost at full well whilst the other is close to zero well. The uncorrected non-linearity is below 2% in many, less exacting situations. In addition, we have examined algorithms that can be used to reduce the maximum residual non-linearity to 3% or less in all cases, with more typical errors being well below 1%. However, reducing the maximum non-linearity even further still may be difficult, and we must await the results of suitable tests with real SBRC array data to fully evaluate the effectiveness of the algorithms discussed here.

Finally, it is worth noting that all the data simulated in this section assumed a normally doped array, for which the variable junction capacitance is quite large. Lower doped arrays, with their correspondingly lower junction capacitance, will have even lower values of non-linearity, probably by a factor of two less than those discussed in this section.

3.8 Noise

We have derived a model for the discharge of a single detector, and have discussed the various parameters that are relevant to the optimum use and calibration of an array of these detectors. Next we must tackle the question of noise, which breaks into two logical sections. The first of these continues the approach thus far, and considers the noise processes at the level of the single pixel, i.e. the uncertainties that arise when we watch one pixel discharging. Secondly, we recognise that we are dealing with not one such detector, but a whole array of them, and that we must tackle uncertainties that arise due to differences between individual pixels in the array, due to partly to manufacturing and processing inhomogeneities.

3.8.1 Noise for an individual pixel

The sources of noise in AC-coupled, non-integrating infrared detectors have been covered at length in the literature (see [51] for example). Such analyses are generally oriented to circuitry employing a photovoltaic diode coupled to a feedback resistor or trans-impedance amplifier, in which the detector operates near zero bias, and the open circuit voltage due to photon illumination is monitored across the feedback resistor at a fairly high frequency. The variations in this voltage are considered as a function of frequency and bandwidth, and various well known figures of merit, such as the noise equivalent power, as thus derived. The four main sources of noise that are considered in such analyses are :

- shot (Poisson) noise associated with statistical variations in the generation and drift of dark current carriers across the p-n junction
- shot noise associated with statistical variations in the arrival rates of source and background photons, and thus with the rate of generation of photo-current
- Johnson noise associated with resistances in parallel with the detector — note that there is no Johnson noise due to the non-ohmic dynamic impedance R_0 of the detector itself
- $1/f$ noise — that is, noise which tends to increase as the sampling frequency is lowered — it is largely dependent on the detector manufacture process, and is not

well understood

However, the classical approach to rigorous noise analysis is not of too much assistance when we come to integrating detectors. The effective sampling frequency is often widely variable, ranging from a few tens of Hz at thermal wavelengths, through to perhaps hundredths of a Hz in low background applications, and figures such as the NEP, usually normalised to 1 Hz, can become somewhat misleading, as they are highly dependent on the nature of the actual application required of a detector. The integrating nature of such detectors tends to smear out the Johnson noise due to stray resistances, and $1/f$ noise components are often decreased through the DC stability of the detector circuit. We are still however involved in a counting process, whether it be of dark current charge carriers, or those generated by photons, and the shot noise associated with them remains of importance. We also introduce the additional component that arises each time we read and reset the detector, the read noise.

3.8.1.1 Read noise

The read noise (σ_{rn}) is defined as the RMS deviation on a signal read out of a pixel after zero integration time. That is, measuring and resetting the pixel many times, with a short an interval as possible between resets, there should be no contribution from dark or photo-generated charge carriers. The signal should therefore be zero, although in practice, some system offset is usually included. In either case, this signal will have an associated RMS noise, and this is the read noise. The read noise is generally quoted in RMS electrons, but is actually measured as a noise voltage, and is converted to electron units via the output circuit capacitance.

It is worth noting that the measurement of read noise is subtly different in practice for an array of the SBRC type, compared to the way it is evaluated for optical CCDs. A charge coupled device clocks the charge collected under each individual pixel out through a single charge sensitive amplifier at one corner of the array. In the rapid read, zero signal situation, the read noise is defined as the RMS deviation in the sensed output voltage as measured for a large number of pixels. In arrays of the SBRC type, each pixel is bump-bonded to its own source follower MOSFET below, namely those in the CRC228 multiplexer for the SBRC array. This is essentially an amplifier of gain just less

than one, which gives a direct voltage measure of the charge on a particular pixel, when it is accessed via the X-Y shift register multiplexer. If we measured the read noise in a fashion analogous to that for a CCD, that is as the RMS deviation from the zero signal voltage for a large number of pixels, then we would in fact be measuring the variation in gains of the individual MOSFETs across the array. This noise will in general be much larger than the actual read noise of any given MOSFET, due to inhomogeneities in the manufacturing process. Thus the true read noise is evaluated by making a series of very short blanked-off exposures, and taking the RMS deviation on the output signal for each detector individually. In practice, although the MOSFET gains in an SBRC array may differ substantially, the read noise is quite similar for the great majority of the pixels, notwithstanding the occasional 'noisy' pixel. This method should also be used for CCDs also, in order to define the true read noise of a single pixel [92].

Whilst state of the art optical CCDs have demonstrated read noises as low as 5–10 e^- RMS [92], integrating infrared arrays have some way to go yet in matching such figures. The read noise of the SBRC 32 × 32 InSb + Si CCD array used by the University of Rochester group has been given as $\leq 1500\,e^-$ RMS [38], and this is typical of the array detectors previously available [106,23]. One of the problems in using silicon CCDs to read out a hybrid array is that the CCD generally needs to be a surface channel device in order to be able to inject the charge generated in the detector substrate into the CCD wells. Surface channel CCDs are noisy, and optical CCD detectors are usually of a backside illuminated, buried channel design to minimise noise due to surface trapping effects. The SBRC 62 × 58 InSb + CRC228 array is newer technology however, and recognising the importance of read noise as a critical parameter to astronomers dealing low photon backgrounds, the array was specified to have a noise floor of 100 μV RMS, which translates to 625 e^- RMS, assuming a typical detector capacitance of 1 pF. Fowler *et al.* [42] found a noise of 77 μV RMS for an SBRC array (which they translated to 575 e^- assuming a capacitance of 1.2 pF), and which they concluded arose in the InSb detector rather than the silicon readout. However, as this noise appears not to be a function of detector bias or gate bias, it may be associated with the integrating mode of the DRO (Hoffman, private communication). Measurements made of lower doped SBRC arrays used in IRCAM have shown read noises as low as 375 e^- RMS.

Finally, the read noise is generally considered as a noise floor, independent of the integration time. As mentioned above, the dominant read noise may arise in the CRC228

circuitry, and it is felt that the CRC228 read noise may be due to $1/f$ noise in the MOSFET [104, page 2] and there is some limited evidence that the read noise may increase for integration times above one second (McLean, private communication). Nevertheless, it is obviously a poorly understood phenomenon as yet, and we are unable to model it at this point. Thus we henceforth assume that the read noise constitutes a noise floor, constant with integration time.

3.8.1.2 Shot noise

The other major source of noise in an integrating detector is the statistical noise due to variations in the rate of generation of charge carriers, due to both dark current and photo-current, the so-called 'shot noise'. A Poisson distribution in arrival rates is generally assumed for such 'counting' experiments, such that, if N individual events are registered, the variance $\sigma^2 = N$.

If we integrate on-chip for a time t, we will in general have collected charge carriers due to the source flux (ϕ_s), the background flux (ϕ_b), and the dark current (I_d), each with an independent shot noise given by :

$$\begin{aligned} \sigma_s^2 &= N_s = \eta\phi_s t \\ \sigma_b^2 &= N_b = \eta\phi_b t \\ \sigma_d^2 &= N_d = I_d t/q \end{aligned}$$

where N_s is the number of source charge carriers collected

N_b is the number of background charge carriers collected

N_d is the number of dark current charge carriers collected

η is the quantum efficiency

q is the electronic charge

Note that it is the total number of charge carriers generated by each source that is important, and not the rate of generation — for example, changes in the instantaneous dark current are not important, as long as the total number of charge carriers generated over the integration time is known.

As the noise contributions are independent, they add quadratically to give the total

shot noise :

$$\begin{aligned} \sigma^2_{shot} &= \sigma^2_s + \sigma^2_b + \sigma^2_d \\ &= \eta\phi_s t + \eta\phi_b t + I_d t/q \end{aligned} \tag{3.78}$$

3.8.2 Spatial noise

Another critical noise component needs to be considered when discussing arrays of simultaneously integrating detectors. This is a noise due to pixel-to-pixel non-uniformity, and is referred to as spatial noise. The main causes of spatial non-uniformity are probably due to defects in the detector crystal structure, and slight variations in the mask used in the photo-lithography process to create the array, generating pixels of slightly different size and quality. Adjacent pixels may individually be background limited, but slight differences in quantum efficiency and readout gain between the two pixels could mean that the background limited performance is not realised, as the theoretical contrast limit between the pixels is masked by pixel-to-pixel variations across the image. In astronomy, we are willing to correct for pixel-to-pixel variations in sensitivity — the flat fielding process — but even after such corrections, residual spatial noise may be present, still limiting the effective sensitivity.

In order to examine the effects of spatial non-uniformity, we shall derive an expression for the spatial noise that may be added to the noise components discussed for an individual pixel. Then we shall study the methods used to remove spatial noise, and the degree to which they may be successful.

For CCD-type detectors, the only spatial non-uniformity usually considered is the pixel-to-pixel variation in quantum efficiency, η. However, the SFD read-out mechanism employed by the SBRC array means we should also consider variations in pixel-to-pixel MOSFET gain, G. Non-uniformity in this term is generally greater than in the quantum efficiency for the SBRC array, and thus must be considered.

Adapting the analysis of Mooney and Dereniak [105], we proceed as follows. Illuminate an array of N pixels with a uniform flux ϕ, and consider the variance, σ^2_{sp}, in the equivalent number of photo-electrons output at the MOSFETs after an integration

time t :

$$\begin{aligned}
\sigma_{sp}^2 &= \frac{1}{N}\sum_{i,j}^{N}\left(G_{ij}\eta_{ij}\phi t - \overline{G\eta\phi t}\right)^2 \\
&= \left(\overline{(G\eta\phi t)^2} - (\overline{G\eta\phi t})^2\right) \\
&= \overline{\phi}^2 t^2\left(\overline{(G\eta)^2} - (\overline{G\eta})^2\right) \\
&= \overline{\phi}^2 t^2 \sigma_{G\eta}^2
\end{aligned} \tag{3.79}$$

where η_{ij} is the quantum efficiency of the i,j^{th} diode

G_{ij} is the electronic gain of the i,j^{th} diode

$\sigma_{G\eta}^2$ is the variance in the product of the quantum efficiency and gain for each diode across the array

Even though non-uniformities in the quantum efficiency and the MOSFET gain are probably uncorrelated, the former arising in the InSb detector material and the latter in the silicon multiplexer, in practice it is difficult to separate the two effects. Thus, in characterising a hybrid array of this type under astronomical conditions, it is the $G\eta$ product that determines the total output spatial variations, and the combined variance σ_{sp}^2 is the parameter we shall deal with.

The most important point to realise about spatial noise that takes the form shown in equation 3.79 is that it can be the limiting factor in determining the ultimate signal to noise for a system. Spatial noise relates to the contrast between adjacent pixels; thus when we have a uniform background falling on the array and it generates spatial noise, we are interested how much additional flux is required to fall on a given pixel such that it can be distinguished from its neighbours. If the mean background flux is ϕ_b, and the additional flux is ϕ_s, then the extra signal falling on the pixel of interest is roughly $\overline{G\eta}\phi_s t$, and the signal to noise is given by :

$$\begin{aligned}
\frac{S}{N} &= \frac{\overline{G\eta}\phi_s t}{\sqrt{\phi_b^2 t^2 \sigma_{G\eta}^2}} \\
&= \frac{1}{U}\frac{\phi_s}{\phi_b}
\end{aligned} \tag{3.80}$$

where U is the non-uniformity, defined as $\sigma_{sp}/\overline{G\eta}$. We can immediately see that for S/N = 1, $\phi_s = U\phi_b$, i.e. a fixed fraction of the sky brightness. Thus, in any situation where spatial noise is dominant, the faintest object distinguishable is defined only by the uniformity of the array and the intensity of the background; this limit is *independent* of

the integration time. Integrating for a longer period will not result in an increased signal to noise for any source. The only way to reach fainter objects or increase the signal to noise on any given object is to reduce the value of U, the non-uniformity. This implies either an intrinsically higher uniformity device (e.g. PtSi Schottky barrier detectors) and/or accurate flat fielding techniques which will reduce the residual non-uniformity, hopefully to a level at which other more fundamental noise sources dominate.

3.8.3 Total noise

Now we can combine the effects of read noise, shot noise, and spatial noise so that we may examine the overall noise for an array of integrating detectors. In order that all the noise sources are referred to a common point, we recall that spatial noise includes the effect of variations in the MOSFET gain across the array, and that the spatial noise was calculated post-MOSFET. The other noise sources were calculated pre-MOSFET, and the variance of each of these sources must be multiplied by a factor of G^2 to normalise them post-MOSFET. In this way, the total variance σ^2_{total} is determined[2] :

$$\sigma^2_{\text{total}} = G^2(\sigma^2_{rn} + \eta\phi_s t + \eta\phi_b t + I_d t/q + U^2\eta^2\phi_b^2 t^2) \tag{3.81}$$

Note that this variance defines the noise for a source of brightness ϕ_s sitting on a background ϕ_b, with mean dark current I_d, read noise σ_{rn}, and with a spatial non-uniformity U, in a *raw* frame. This is not a particularly helpful way of characterising the noise of real interest, namely the noise left after the data have been reduced. It is this noise which determines the actual astronomical sensitivity of the system.

We shall use a reduction technique similar to that outlined in section 3.6.1 to derive the final noise. Even though this method assumes a linear detector, the actual steps performed are very similar to those used when linearity corrections are included (see section 3.7), and the propagation of the noise sources will be substantially similar.

Firstly we drop the spatial noise term from the analysis. Whilst this may seem strange after having decided that spatial noise may be very important, we must realise

[2]Note that we no longer refer explicitly to the mean quantum efficiency and mean gain across the array—this is implicit and is to prevent cluttering the following equations. Since we have now quantified the spatial noise in terms of U, replacing the actual values of G and η with their means will be valid for any given pixel as long as U is not too great (a few percent)

that we do not understand the way in which the errors due to non-uniformity propagate. That is; if a perfect flat field were used to correct for the pixel-to-pixel sensitivity variations, there would be no residual non-uniformity left, and the remaining noise on the data would be due the other effects alone. However, as discussed in section 3.10, some residual non-uniformity generally remains in practice. We shall reintroduce the spatial non-uniformity at the end of the analysis in the form of U_r, the residual non-uniformity.

Assume a fixed on-chip integration time of t for every exposure. At this stage, we will analyse the case of one co-added exposure only. We shall see later how the noise may be reduced by co-adding many dark, flat-field, and source exposures. Assume the system offset O to be fixed and similar for every pixel. The dark signal S_D in a given pixel and its variance σ_D^2 are :

$$S_D = O + GI_d t/q \tag{3.82}$$

$$\sigma_D^2 = G^2\left(I_d t/q + \sigma_{rn}^2\right) \tag{3.83}$$

The signal S_F and variance σ_F^2 for the flat field frame, taken on a piece of blank sky near the source field, is similarly defined :

$$S_F = O + Gt(I_d/q + \eta\phi_b) \tag{3.84}$$

$$\sigma_F^2 = G^2\left(I_d t/q + \sigma_{rn}^2 + \eta\phi_b t\right) \tag{3.85}$$

Similarly for the signal S_S and variance σ_S^2 in the source frame :

$$S_S = O + Gt(I_d/q + \eta(\phi_b + \phi_s)) \tag{3.86}$$

$$\sigma_S^2 = G^2\left(I_d t/q + \sigma_{rn}^2 + \eta\phi_b t + \eta\phi_s t\right) \tag{3.87}$$

Now, following the procedure of section 3.6.1, we subtract the dark frame from both the flat field and source frames, noting that the errors simply add in quadrature at this stage :

$$S_F - S_D = Gt\eta\phi_b \tag{3.88}$$

$$\sigma_{(F-D)}^2 = G^2\left(2I_d t/q + 2\sigma_{rn}^2 + \eta\phi_b t\right) \tag{3.89}$$

$$S_S - S_D = Gt\eta(\phi_b + \phi_s) \tag{3.90}$$

$$\sigma_{(S-D)}^2 = G^2\left(2I_d t/q + 2\sigma_{rn}^2 + \eta\phi_b t + \eta\phi_s t\right) \tag{3.91}$$

Next we divide the dark subtracted source frame by the dark subtracted flat field frame. Then we rescale the result by the mean of the dark subtracted flat field frame. Lastly,

we subtract the mean sky value from the data. This may be obtained from the image frame itself, or from the flat field frame, which was taken on a nearby patch of sky. We assume the latter case here. This leaves us with a rescaled and sky subtracted image, which is what we want. Note that the errors add in a slightly more complex fashion here, as we are dividing one frame by another. The rescaling and sky subtracting processes are assumed to be noiseless, as only a single scalar is used in each case, which may be determined from many pixels. The results are :

$$\left(\frac{S_S - S_D}{S_F - S_D} \times \overline{S_F - S_D}\right) - \overline{S_F - S_D} = Gt\eta\phi_s \tag{3.92}$$

$$\sigma^2 = G^2\left(2I_dt/q + 2\sigma_{rn}^2 + \eta t(\phi_b + \phi_s) + \left[\frac{\phi_s + \phi_b}{\phi_b}\right]^2\left(2I_dt/q + 2\sigma_{rn}^2 + \eta\phi_b t\right)\right) \tag{3.93}$$

It is here that we reintroduce the spatial noise in the form of U_r, the *residual* spatial non-uniformity. The variance resulting from this is given by analogy to equation 3.79 :

$$\sigma_{sp}^2 = U_r^2 G^2 \eta^2 \phi_b^2 t^2 \tag{3.94}$$

Adding this term into the noise expression, we can now write down an expression for the signal to noise obtained for a source of flux ϕ_s on a background ϕ_b, with dark current I_d and read noise σ_{rn}, in this case *after reduction* :

$$\begin{aligned} \text{Signal} &= G\eta\phi_s t \\ \text{Noise} &= G\Bigg[\, U_r^2\eta^2\phi_b^2t^2 + \left(2I_dt/q + 2\sigma_{rn}^2 + \eta t(\phi_b + \phi_s)\right) \\ &\qquad + \left[\frac{\phi_s + \phi_b}{\phi_b}\right]^2\left(2I_dt/q + 2\sigma_{rn}^2 + \eta\phi_b t\right)\,\Bigg]^{1/2} \end{aligned} \tag{3.95}$$

Before looking at this equation and discussing its component parts, we shall introduce the final variable, as promised earlier : this is the number of co-added exposures that go to make up each output frame. That is, in order to reduce the effects of read noise and temporal noise, many on-chip exposures may be co-added in computer memory. The number of co-adds may be different for the source, flat field, and dark current measurements — the 'calibration' frames may have many more co-adds than the source frame in order to reduce the extra noise added during the reduction procedure. For example, co-adding N dark current exposures should reduce the shot and read noise components in the mean dark current frame by a factor of $\sqrt{N}$, thus reducing the noise added by the dark subtraction process. If we denote the number of co-added exposures that go to make up the final source, flat field, and dark current frames by N_S, N_F, and

N_D respectively, we can rework the previous analysis, and arrive at a modified version of the previous equation accordingly :

$$
\begin{aligned}
\text{Signal} &= N_S G \eta \phi_s t \\
\text{Noise} &= G \Bigg[N_S^2 U_r^2 \eta^2 \phi_b^2 t^2 + && \text{(a)} \\
&\quad \left(N_S (I_d t/q + \sigma_{rn}^2 + \eta t(\phi_b + \phi_s)) \right) + && \text{(b)} \\
&\quad \left(\left(\frac{N_S^2}{N_D} \right) (I_d t/q + \sigma_{rn}^2) \right) + && \text{(c)} \\
&\quad \left(\left[\frac{\phi_s + \phi_b}{\phi_b} \right]^2 \left(\frac{N_S^2}{N_F} \right) (I_d t/q + \sigma_{rn}^2 + \eta \phi_b t) \right) + && \text{(d)} \\
&\quad \left(\left[\frac{\phi_s + \phi_b}{\phi_b} \right]^2 \left(\frac{N_S^2}{N_D} \right) (I_d t/q + \sigma_{rn}^2) \right) \Bigg]^{1/2} && \text{(e)}
\end{aligned}
\tag{3.96}
$$

Although this equation may appear somewhat daunting, it is in fact quite simple. The source of the noise components inside the $G[\cdots]^{1/2}$ can be explained as follows :

- component (a) is the residual non-uniformity term — note that it has a factor N_S^2 in it, implying that in a regime dominated by spatial noise, the signal to noise is independent of the number of co-added frames, similar to the way that it was seen to be independent of integration time in section 3.8.2.

- component (b) is due to the shot noise and read noise in the raw source frame itself.

- component (c) is due to the noise in the dark current frame which was subtracted from the source frame, scaled according to the number of exposures co-added — note that for $N_D = N_S$, this term adds with equal weight to term (b), but if many dark current exposures are made ($N_D \gg N_S$), then the noise added by this component is much reduced.

- component (d) is due to the noise in the raw flat field frame. Note the two scaling factors — the term $[(\phi_s + \phi_b)/\phi_b]^2$ arises when the source frame is divided by the flat field frame, the result being rescaled by multiplying by the mean value in the flat field frame. This factor shows that for faint sources ($\phi_s \ll \phi_b$), component (d) contributes just as much noise as component (b), and that when $\phi_s \gg \phi_b$, the noise from component (d) will *dominate* that from component (b). That is; for

bright sources, where you might expect the signal to noise to be limited by shot noise on the source itself, it is in fact limited by the shot noise on the background, i.e. the noise in the flat field. Thus, in order to increase the signal to noise on the source frame in this case, the other factor (N_S^2/N_F) must be made smaller; that is, N_F must be made $\gg N_S$.

- component (e) is due to the noise on the dark current frame subtracted from the flat field frame — the comments attached to components (c) and (d) are relevant here also.

Having derived a general expression for the theoretical signal to noise, we can discuss the ways in which this signal to noise may be modified by changes in the various parameters which go to make it up. To do this, we define a set of regimes in each of which a certain noise source is assumed to be dominant. In the following sections, we assume in each case that the number of co-adds used for each of the dark, flat field, and source frames is the same, for the sake of simplicity. It is evident from equation 3.96 that by taking many more dark and flat field co-adds than source co-adds, the final errors in the reduced source image will be minimised. This point is addressed further where relevant below. The first four sections deal with the case of $\phi_s \ll \phi_b$, i.e. the source flux being a small perturbation on the background, such that we are looking at the sensitivity to faint objects. In each section we examine how the signal to noise depends on the relevant parameters. We examine how the signal to noise changes as a function of the area of a pixel subtended on the sky, the area of the telescope, and how the signal to noise may be improved for faint diffuse sources by adding up the signal over several pixels—the 'binning-up' procedure.

3.8.3.1 Read noise limited

The read noise limited regime occurs when $\sigma_{rn}^2 > I_d t/q, \eta\phi_s t$ and $\eta\phi_b t$, and the residual spatial noise is negligible. In the limit, equation 3.96 contracts to :

$$\begin{aligned} \text{Signal} &= \sqrt{N_S}\,\eta\phi_s t \\ \text{Noise} &= 2\sigma_{rn} \end{aligned} \tag{3.97}$$

This gives us the following dependencies :

S/N increases as the square root of the number of co-adds

increases linearly with the integration time

increases linearly with the source brightness

decreases linearly with the read noise

increases linearly with the area imaged by a pixel

increases linearly with the area of the telescope

increases as the square root of the number of pixels averaged together

For the SBRC array with a typical read noise $\sigma_{rn} \sim 400\,e^-$ RMS, an individual pixel will be read noise limited until the well is filled to $\sim 400^2$ electrons, i.e. until about $\sim 1.6 \times 10^5$ charge carriers have been generated by the dark and/or photo-currents. As the well depth is typically $\sim 10^6\,e^-$, read noise can easily be overcome. By filling to half well before reading out, the read noise becomes negligible, and the device dynamic range remains large. The read noise dominated case may arise where high time resolution is required in speckle observations for example, or if magnitude of the dark and photo-currents are both so small that the integration time taken to fill to half well is impracticably long, as may be the case for very high resolution spectrometers operating in the 1–2.5 μm range.

3.8.3.2 Dark current shot noise limited

This case occurs when $I_d/q > \eta\phi_s$ and $\eta\phi_b$, and with the residual spatial noise negligible. Enough charge carriers must be integrated to overcome the read noise as discussed above. Then equation 3.96 is reduced to :

$$\begin{aligned} \text{Signal} &= \sqrt{N_S t}\,\eta\phi_s \\ \text{Noise} &= 2\,[I_d/q]^{1/2} \end{aligned} \tag{3.98}$$

This gives us the following dependencies :

S/N increases as the square root of the number of co-adds

increases as the square root of the integration time

increases linearly with the source brightness

decreases as the square root of the dark current

increases linearly with the area imaged by a pixel

increases linearly with the area of the telescope

increases as the square root of the number of pixels averaged together

Cooling the SBRC array to 40 K reduces the mean dark current to $\sim 100\,e^-$/sec/pixel. Only in very few imaging applications in the 1–5.5 μm region will the photon background be that low. High resolution spectroscopy between 1 and 2.5 μm may result in background photo-currents as low as $\sim 1\,e^-$/s/pixel, however. Dark currents in InSb are unlikely to fall that low even at 4 K, so in order to avoid being dark current noise limited, larger bandgap semiconductor detectors (e.g. $Hg_{0.554}Cd_{0.446}Te$) should be used.

3.8.3.3 Background shot noise limited

This case arises when $\eta\phi_b > I_d/q$ and $\eta\phi_s$, with the spatial noise negligible, and enough charge carriers integrated to overcome the read noise. Equation 3.96 reduces to :

$$\begin{aligned} \text{Signal} &= \sqrt{N_S \eta t}\,\phi_s \\ \text{Noise} &= 2\,[\phi_b]^{1/2} \end{aligned} \tag{3.99}$$

This gives us the following dependencies :

S/N increases as the square root of the number of co-adds

increases as the square root of the integration time

increases linearly with the source brightness

decreases as the square root of the sky background brightness

increases as the square root of the area imaged by a pixel

increases as the square root of the area of the telescope

increases as the square root of the number of pixels averaged together

This situation will be the one most frequently encountered in imaging applications in the 1–5.5 μm region, as long as the spatial noise remaining after the flat fielding process is negligible. It provides the theoretical limit to faint object sensitivity that can be achieved, background limited performance. Once a system achieves this performance, it is doing as well as any other system can on the same telescope, provided the sky background brightness is as low as possible, and that no high emissivity sources are seen in the beam at thermal wavelengths. If a system is background limited, then the

chosen pixel field of view is irrelevant, as the increase in signal to noise for diffuse sources as the n pixels are added together is $\sqrt{n}$, and the increase in signal to noise as the area A of sky imaged by each pixel is increased is $\sqrt{A}$. As $A \propto n$, there is no advantage in using bigger pixels initially, as the same signal to noise may be achieved by binning up smaller background limited pixels until the desired signal to noise is reached. In fact, as soon as the diffuse emission becomes clumpy, smaller pixels win, as they reach the required signal to noise with fewer pixels in the bright clumps, keeping the flexibility of variable spatial resolution binning across the image.

3.8.3.4 Source shot noise limited

This situation should arise when $\eta\phi_s > I_d/q$ and $\eta\phi_b$, as long as residual spatial noise is negligible, and the read noise is overcome by integrating enough charge carriers before reading out. Note that our assumption of $\phi_s \ll \phi_b$ is no longer true, by definition, and we must be more careful when reducing equation 3.96 to :

$$\begin{aligned} \text{Signal} &= \sqrt{N_S \eta t}\, \phi_s \\ \text{Noise} &= \left[(\phi_b + \phi_s) + \left[\frac{\phi_s + \phi_b}{\phi_b}\right]^2 \phi_b \right]^{1/2} \end{aligned} \tag{3.100}$$

This gives us the following dependencies :

S/N increases as the square root of the number of co-adds

increases as the square root of the integration time

increases as the square root of the area imaged by a pixel

increases as the square root of the area of the telescope

increases as the square root of the number of pixels averaged together

The usual assumption made made for sources much brighter than the background is that the signal to noise increases as the square root of the source brightness; that is, that we are limited by the shot noise on the source photons themselves, the theoretical limit. However, careful examination of equation 3.100 reveals that this is *not* the case for arrays which have to be flat fielded. The first term in the noise $(\phi_b + \phi_s)$ would suggest that, for $\phi_s \gg \phi_b$, the overall signal to noise *would* increase as the square root of the source brightness. But it is the second term that is critical, the term introduced in the flat fielding process. It can be seen that as ϕ_s becomes very much larger than

ϕ_b, the $[(\phi_s + \phi_b)/\phi_b]^2\phi_b$ term increases *as the square of* ϕ_s : that is, the signal to noise achieved tends asymptotically towards a fixed limit. That limit is the set by the background photon shot noise in the flat field frame. Thus, it does not matter how much brighter than the background a source is in the source frame, the signal to noise achieved is set by the signal to noise in the flat field frame. This gives us a clue as to how to improve the overall signal to noise. As stated earlier, co-adding many more flat field and dark exposures than source exposures will reduce the final error. In this case, equation 3.96 shows us that the signal to noise for a bright source is proportional to $\sqrt{N_F}$, where N_F is the number of co-added flat field exposures.

3.8.3.5 Residual spatial noise limited

This is the case where, after flat fielding has taken place, the residual spatial noise dominates the noise due to all the other terms. In this case, equation 3.96 reduces simply to :

$$\begin{aligned} \text{Signal} &= \phi_s \\ \text{Noise} &= U_r\phi_b \end{aligned} \tag{3.101}$$

This gives us the following dependencies :

S/N	does not increase at all with the	number of co-adds
	does not increase at all with the	integration time
	does not increase at all with the	area imaged by a pixel
	does not increase at all with the	area of the telescope
	increases linearly with the	source brightness
	decreases linearly with the	sky background brightness
	increases as the square root of the number of pixels averaged together	

For a given ratio of source to background intensities, the only way to increase the signal to noise (which is independent of integration time, telescope size, pixel size, and number of co-adds) is to decrease the residual spatial non-uniformity U_r. This may be done by binning-up pixels, or more effectively by stacking many images taken with slight field centre offsets, as this does not compromise the spatial resolution. Note however that

the same must be done for the flat field in this case, as it is obvious by analogy to the discussion of section 3.8.3.4 that a high signal to noise frame will be compromised if the flat field frame has lower signal to noise, even though this effect does not appear explicitly in equation 3.96 due to the way in which we dealt with spatial noise in deriving that equation.

3.9 Noise characteristics of the SBRC array

Here we shall discuss the characteristics of noise components seen in an SBRC array when in use in an astronomical situation. We shall do this by reducing some astronomical data, finding the final noise in the sky background, and then inputting the relevant noise components into the model of equation 3.96 to predict the noise that should be seen. This will allow us to decide :

- if the theory developed in section 3.8 is valid for the SBRC array
- the magnitude of the inherent spatial non-uniformity U
- the magnitude of the residual spatial non-uniformity U_r after the data have been processed
- how well the IRCAM flat fielding technique works and the implications for more effective calibration

The data we shall examine were all obtained on the 19th April 1987 with IRCAM on the UKIRT, as part of a photometric calibration experiment. The following system parameters applied :

Detector array	:	FPA039
Read-out rate	:	129.6 msecs
System gain	:	24 e^-/DN
V_{bias}	:	−0.25 V
V_{gate}	:	−3.05 V
T_{det}	:	33.5 K
Pixel size	:	0.6″/pixel

sets

We shall examine three data taken through three different filters, each with a different on-chip integration time and number of co-added exposures. In each case, the number of co-added exposures used to obtain the sky or flat field frame was the same as the number used for the source frame. However, the number of dark current exposures co-added was different in each case, with an important effect on at least one result, as we shall see. Throughout, the data frames are normalised by dividing each by the number of co-adds used. In this way, equation 3.96 is slightly modified by the inclusion of a $1/N_S$ term at the beginning of both the signal and noise expressions. The fact that a different number of co-adds was used for the dark frames is embedded in the main noise expression itself. Throughout, statistics are presented as the data are reduced. The median and standard deviation were evaluated over a clean area of the array seeing only sky; we are not interested in the sources in the image, just the noise on the background. We use the median to be sure that odd bad pixels do not affect the results. However, in every case, the median value was very close to the mean, indicating that no bad pixels were in fact included. This means that the standard deviation is a valid estimate of the true noise, also unaffected by bad pixels.

The numbers we need to derive in each case are :

- the mean dark current
- the mean photo-current
- the final noise on the background actually arrived at after reduction

We assume a read noise of 375 e^- RMS throughout, as determined from repeated measurements of a single pixel with a very short integration time. Errors in this assumption will not be important, as each data set is in fact background limited.

3.9.1 Example 1 : K broad band

The first example looks at data obtained looking through the K broad band filter. The data are :

Frame	Filter	τ_{int}	No. co-adds
Bias	—	0.13 s	10
Dark	—	8 s	10
Sky	K	8 s	13
Source	K	8 s	13

Each frame is divided by the relevant number of co-adds, and statistics are taken in a box in a clean area of size 5 × 15 pixels, lower left corner at pixel 55,34. The results are :

Bias	Median =	34824 DN	σ =	379 DN
Dark		42174		407
Sky		51815		445
Source		51856		447

We assume that the bias frame does represent the zero signal level, as its integration time of 0.13 seconds is negligible compared to the 8 seconds of the dark frame. We subtract the bias frame from the dark frame, and statistics are taken in the same box in the resulting frame :

Dark − Bias	Median =	7403 DN	σ =	192 DN

We multiply by the gain of 24 e^-/DN, and divide by the on-chip integration time of 8 seconds to obtain the mean dark current of 22,209 e^-/second/pixel. This is very high for an operating temperature of 33.5 K, and is probably due to an electroluminescent effect on the detector chip rather than true dark current. However, we can still use this number in the analysis, as it will obey shot noise statistics as does real dark current.

Next we need to find the photo-current. We subtract the dark current frame from both the sky and source frames, noting that this also removes the bias offset. The statistics for the resulting frames are :

Sky − Dark	Median =	9508 DN	σ =	253 DN
Source − Dark		9541		254

It is the photo-current in the source frame we are interested in, even though it is very close to that of the sky frame, as we are taking statistics in a clear region of background

in both. We multiply the median of the selected region in dark subtracted source frame by the gain and divide by the on-chip integration time, to give the mean photo-current of 28,623 e^-/second/pixel. Note that this is about 2.5 times higher than expected for this imaging mode at K (see chapter 4).

Next we flat field the data. We divide the dark subtracted source frame by the dark subtracted sky frame. Then we re-scale, multiplying by the median value found over the defined box in the sky frame. Statistics in the resulting flat fielded frame give :

Flat fielded source Median = 9540 DN σ = 16.0 DN

Note how the noise was dramatically reduced in the flat fielding process. The flat field is obviously very good, as the result constitutes a noise of $\sim 0.2\%$ of the sky background, better than the usual CCD result of 1%.

Next we can use the figures derived in the analysis to predict the noise figure from the model. We can then compare the result with the noise of 16.0 DN actually obtained. In this case, where we have only background flux, $N_S = N_F$, $N_D \neq N_S$, and where we have normalised each frame by dividing by the number of co-adds, equation 3.96 reduces to :

$$\text{Noise} = \frac{G}{N_S}\Big[N_S^2 U_r^2 \eta^2 \phi_b^2 t^2 + 2N_S(I_d t/q + \sigma_{rn}^2 + \eta t \phi_b) + \frac{2N_S^2}{N_D}(I_d t/q + \sigma_{rn}^2) \Big]^{1/2} \quad (3.102)$$

We use a read noise of 375 e^- RMS, a mean dark current of 22,209 e^-/sec/pixel, a mean background photo-current of 28,623 e^-/sec/pixel, $N_S = 13$, $N_D = 10$, and the gain of 24 e^-/DN (which implies $G = 1/24$ DN/ e^-). As, by definition, we do not yet have a value for U_r, we assume it is zero, evaluate the noise expression, and then attribute any excess noise to U_r. Putting the numbers into the noise equation gives a predicted noise figure of 16.03 DN.

This is the same result as was found after reducing the actual data. Thus, any contribution from residual spatial non-uniformity must be very small indeed. In order to increase the predicted noise result from 16 DN to 17 DN, the parameter U_r would need to be $\sim 6 \times 10^{-4}$. Therefore, it probably fair to say that we have reduced the residual

spatial non-uniformity to $\leq 0.05\%$ in this data set, and that we are definitely limited by the shot and read noise components. However, before drawing any conclusions, we shall reduce the other two data sets, in an attempt to confirm the result presented here.

3.9.2 Example 2 : J broad band

The data used in this example are :

Frame	Filter	τ_{int}	No. co-adds
Bias	—	0.13 s	10
Dark	—	4 s	10
Sky	J	4 s	50
Source	J	4 s	50

We follow exactly the same reduction procedure as for the first example. The statistics box used this time is a 5 × 15 pixel box, lower left corner at pixel 46,34. The statistics for the normalised frames are :

Bias	Median = 34748 DN	σ = 306 DN
Dark	38643	333
Sky	58415	389
Source	58491	392

The bias frame is subtracted from the dark current frame. The statistics for the result are :

Dark − Bias	Median = 3887 DN	σ = 152 DN

This gives a mean dark current of 23,319 e^-/second/pixel. This is slightly higher than for the previous example, probably due to a combination of taking a different box in this example, and the non-linearity of the dark current, which causes it to decrease as the diode is discharged — as this data has an on-chip integration time of only 4 seconds, the mean dark current will be slightly higher than in the previous example, where the on-chip integration time was 8 seconds. Next, we subtract the dark current frame from the sky and source frames. The statistics in these frames are :

Sky − Dark	Median =	19839 DN	σ =	333 DN
Source − Dark		19912		335

The mean photo-current is $1.1947 \times 10^5\,e^-$/second/pixel. This is about a factor of 100 times higher than it should be at J, and is due to a red leak in the J filter (see chapter 4).

Next we flat field as before. The statistics in the renormalised result are :

Flat fielded source	Median =	19907 DN	σ =	12.3 DN

This is a very low noise, corresponding to 0.06% of the sky background level, close to the best ever obtained with optical CCD detectors. Using a read noise of $375\,e^-$ RMS, a mean dark current of $23{,}319\,e^-$/sec/pixel, a mean photo-current of $1.1947 \times 10^5\,e^-$/sec/pixel, $N_S = N_F = 50$, $N_D = 10$, $G = 1/24$ DN/e^-, and again assuming U_r is zero until proved otherwise, we calculate the predicted noise from equation 3.103, arriving at a result of 11.44 DN. This is very close to the actual noise figure again, within 1 DN. In fact, examining the flat fielded data frame another area of the array was found to be clean. This area was larger, containing three times as many pixels as the control box — this larger box had a final noise of 11.51 DN, even closer still to the predicted noise of 11.44 DN. Therefore, it is true again that any residual spatial non-uniformities left after flat fielding are very small, and that the noise performance in this example is also completely explained by shot and read noise alone.

3.9.3 Example 3 : narrow band L

In this example, we choose data with a much higher background rate, and with many more co-added frames. The data used were taken through the narrow band L filter at 3.6 μm, and are :

Frame	Filter	τ_{int}	No. co-adds
Dark	—	0.13 s	10
Sky	nbL	0.13 s	1000
Source	nbL	0.13 s	1000

Note that we do not have a bias frame this time. This is because an on-chip integration time of 0.13 seconds was used for the dark current, sky, and source exposures, and thus

a bias frame would be just the same as the dark frame. The statistics box used this time is a 5 × 15 pixel box, lower left corner at pixel 55,34. The statistics for the normalised frames are :

Dark	Median =	34835 DN	σ =	337 DN
Sky		46139		259
Source		46185		259

In this example, we simply use the mean dark current derived in the second example, i.e. 23,319 e^-/sec/pixel. As the amount of dark current integrated in 0.13 seconds is very small (i.e. 3,031 e^-/pixel), an error in this assumption will not be important, as the noise in the dark current frame will be heavily dominated by read noise. Next we subtract the dark current frame from both the sky and source frames, with the statistics for the resulting frames as follows :

Sky − Dark	Median =	11324 DN	σ =	323 DN
Source − Dark		11366		323

This gives us a mean photo-current in the source frame of 2.0983×10^6 e^-/sec/pixel. This is approximately 50 times higher than predicted for this filter (see chapter 4). Dividing and re-scaling, we flat field the source frame, the statistics for the result being :

Flat fielded source	Median =	11367 DN	σ =	6.0 DN

Again, this is extremely good flat fielding accuracy, ~ 0.05% of the sky background. We now calculate the predicted noise value for these data. Taking a read noise of 375 e^- RMS, a mean dark current of 23,319 e^-/sec/pixel, a mean photo-current of 2.0983×10^6 e^-/sec/pixel, $N_S = N_F = 1000$, $N_D = 10$, $G = 1/24$ DN/e^-, and setting U_r to zero initially, we arrive at a predicted noise of 7.2 DN from equation 3.103. This is slightly higher than the actual noise result, but still close. That the actual data should have *less* noise after reduction than the theoretical prediction is a surprise. At least it again implies that there is no contribution from residual spatial non-uniformity, but quite why the result is less noisy is unknown. In fact, in another clean part of the array, the noise falls as low as 4 DN. Note however that the dominant source of noise in the reduction process is the read noise in the dark current frame. This is because the

number of co-adds used in the dark current frame was much smaller than used in the sky and source frames, and the N_S^2/N_D term became much larger than N_S in the noise equation. Thus the term due to dark subtraction dominates the noise equation, and as the shot noise on the dark current accumulated in 0.13 seconds was small relative to the read noise, the read noise dominated the final result, even though the source and sky frames were heavily shot noise limited. If N_D had been 1000 also, the predicted noise would have been significantly reduced, to about 1.4 DN. Thus, an extra two minutes integration time (1000×0.13s) spent integrating on a dark frame (which constitutes an overall increase in time spent of 50%, or less if the overheads involved in starting a frame of any length are accounted for) would have brought a final signal to noise gain of approximately five.

3.9.4 Conclusions from noise examples

From the data analysed in the three examples above, we can draw the following conclusions about the behaviour of noise in fully reduced data from the SBRC array, and about the predictive powers of the model detailed in equation 3.96 :

- The noise obtained over a range of astronomical conditions is fully consistent with theoretical predictions based on read noise, and dark current and photo-current shot noise components alone.

- The residual spatial non-uniformity (U_r) appears to be completely negligible in the situations discussed above, at a level of $\ll 0.05\%$ of the sky background.

- No additional noise mechanisms appear to be necessary to explain the noise obtained in the actual data. This implies that $1/f$ components in the either the sky or detector are minimal, at least on the timescales detailed above ($\sim$ several minutes).

- The flat fielding process used seems highly satisfactory, reducing all systematic errors well below the noise levels expected from read and shot noise.

- Care must be taken that the noise after reduction is not compromised by insufficient signal to noise in the calibration frames. This means that at least the same number of co-adds should be used for dark and sky frames as used for the source

frame itself, and that additional gains in signal to noise would be possible if the number of co-adds used for the calibration frames were greater than those used for the source frames.

- With a relatively stable and low dark current, it should be possible to take a whole series of frames with a range of integration times and each with a large number of co-adds, before observing begins. Spot checks on the dark current during the night will indicate whether or not it is drifting.

- The same procedure would prove difficult for sky frames used as flat fields, as the extremely accurate flat fielding no doubt owes much to the fact that the flat field is an exposure taken on the sky nearby spatially and temporally to the actual source frame, as discussed in section 3.10. Increasing the number of co-adds used for sky frames over the number of co-adds used for the source frames would be of diminishing benefit, as the signal to noise in the source frame becomes dominant, at the expense of wasted on-source integration time. Thus, keeping the numbers of co-adds equal for sky and source frames is recommended as a simple compromise.

- The SBRC array appears to be relatively free of unexpected noise components. As it is the most sensitive array of its type, and the UKIRT is the largest dedicated infrared telescope in the world, on the best site, the results that may be achieved with their combination would appear to be unrivalled at this time.

Finally, we note that little has been said about the inherent spatial non-uniformity of the SBRC array, that is, the pixel-to-pixel variations that exist in raw data frames. This is because at this time, it is thought that the noise seen in raw data frames is probably due to the electronic processing of the output data, rather than inherent non-uniformities. The limited evidence for this stems from the fact that even after the very shortest integration time, there is a large noise figure. If the array is reset and read out very rapidly in the dark, each pixel should be at the same voltage it started at, no discharge having taken place. The sample and reset voltages will be equal for each pixel, and even though these voltages may be different across the array due to real gain differences from pixel-to-pixel, the result output by the processing electronics should be zero in all cases, leading to a noise free image. The fact that this is not true leads us to believe that the noise may be due to insufficient common mode rejection at the twin sample and hold (SH) units that form the difference signal between the sample and

reset voltages. The two voltages arriving at the SH units sit on a pedestal voltage, the magnitude of which is a function of the gain of the MOSFET of each pixel and therefore differs across the array. If the SH units are incorrectly balanced, then they will allow a fraction of the pedestal voltage through, despite both SH units being presented with the same voltage. As the pedestal varies across the array, the voltage allowed through the SH units will vary, causing a noise. This noise should disappear in the data reduction process, as it seems to do, but makes it very difficult to derive the inherent pixel-to-pixel non-uniformities. Data from SBRC would indicate that the value of the inherent spatial non-uniformity U, is on the order of 5% [116].

3.10 Flat fielding

The SBRC arrays in IRCAM do not appear to be limited in any way by spatial non-uniformity. We shall now discuss how residual non-uniformities may still persist after flat fielding, and how their effects may be reduced in practice, thus explaining why the method used in IRCAM is so good.

3.10.1 Flat fielding optical CCDs

Once a detector is manufactured, the inherent spatial non-uniformity U is defined. The flat fielding technique assumes that by illuminating the detector with a uniform flux, a template of the pixel-to-pixel gain differences can be obtained, and subsequent exposures on sources of interest can be divided by this template, in order to normalise out the gain differences. We shall examine the ways in which optical astronomers employ this technique, beginning from the raw method, which uses diffusely illuminated exposures of the inside of the dome to provide a flat field.

- Using the dome lights (usually tungsten at $\sim 3000\,\mathrm{K}$) leads to differences between the colour temperatures of the source and flat field fluxes This is important, as the device quantum efficiency variations are often wavelength dependent

- Filtering the flat field source may provide the correct colour temperature, but still provides a poor analogue to the source flux, which generally consists of sources

with a range of black-body temperature spectra on top of a night sky background mainly of narrow emission lines

The sky may be used as a flat field source, but for optical CCDs there are the following disadvantages.

- CCD detectors have quite large area coverage, and it is impossible to find a blank piece of sky. Several sky flats must be taken in order to remove stars and other objects from the field.

- The sky may be quite faint in the optical, and thus it takes a long time to get a high enough signal to noise flat field that it will not significantly degrade the signal to noise of the source frame.

- The best flat fielding is done when the intensity of the flat field is similar to the intensity of the source being flat fielded, due to non-linearities in the detector. Thus, whilst faint objects near the sky background are well catered for with sky flats, brighter objects are not. Two different illumination levels may be required to bracket the range of object brightnesses, and this may take a long time, integrating on sky alone.

- During the long integration time typical of CCD exposures, the spectrum and intensity of the night sky may change dramatically, and a sky flat obtained either before or after the source frame may be inadequate.

- Lastly, there are many monochromatic emission lines in the night sky spectrum at the red end of the optical. Thinned backside illuminated devices have a detector substrate just deep enough to allow detection of the longest wavelengths, which penetrate deep before being detected, but not so deep that the electron-hole pairs generated close to the surface by the more energetic short wavelength photons recombine before diffusing to the p-n junction. Thus, the material is almost transparent to the long wavelength photons. Thinned devices are often mounted on glass plates, and photons which penetrate the detector may be reflected at the detector-glass interface, where a large change in refractive index occurs. This promotes multipath interference effects in a process as 'fringing'. The fringes created may modulate the imaged background by up to 100%, and are very difficult to remove in the flat fielding process.

Two groups using optical CCDs have consistently obtained lower residual spatial non-uniformities than the usual limit of 1%, and it is worth briefly examining their techniques in order to determine the best way to flat field an infrared array.

3.10.1.1 Drift scan

The Cambridge group use the 'drift scan' technique [164]. This involves taking images with the telescope right ascension drive slowed or turned off, and with the CCD columns aligned exactly east-west. As the astronomical scene drifts through the focal plane in a westerly direction, the charge packets stored under each pixel are slowly clocked out along the columns of the array at such a rate that a given accumulated charge packet effectively stays beneath the same point in the astronomical scene. On reaching the end of the column, the charge packet is read out as normal, and a continuous strip scan image is built up in the computer. In this way, the pixel-to-pixel variations along the length of a column are eliminated. Then the column-to-column sensitivity variations are removed by comparing the median value of each column. Combined with a technique that tries to account for large scale variations in the sky background, it is possible to realise background limited performance with this technique, as opposed to the spatial noise limited performance usually achieved—a limit of 0.1% of the sky background is quoted as the residual non-uniformity [58].

3.10.1.2 Adaptive modal filtering

The SBRC array does not have a CCD-type readout, but rather an X-Y shift register addressing scheme. Thus the drift scan technique is not viable. The second technique is however. It is known as 'adaptive modal filtering', and is due to Tyson [146]. This technique is less sophisticated in terms of the actual observations, but is much more sophisticated in the data analysis. Many frames are taken of an object field, with each image offset from the last by small and random amounts. Flat field frames are made by looking at a diffusely illuminated surface through the required filters. To generate the master source frame, all the individual frames are flat fielded, and then registered up on a bright point source and stacked. The final value for each pixel in the image (x,y) plane is determined by forming the mode of all the values for that given point in space in the

stack (z) plane. Taking the mode eliminates the fringes due to the lines in the night sky, as they stay fixed on the array as the objects move around. Bad pixels and cosmic ray hits are also removed in this way. The residual spatial non-uniformity is reduced in this way by the square root of the number of stacked images, as are all the read and shot noise components. Whilst this is the basic principle, the software is more sophisticated than described : setting various threshold parameters, the process is iterated upon several times, and residual non-uniformities of 0.03% of the sky background have been reported, at which point the system is background limited [146].

3.10.2 Flat fielding the SBRC array

The default method of flat fielding the SBRC array in IRCAM is as follows. In a DC analogue to chopping, images with the same integration time and number of co-adds are made of a piece of blank sky close to the object of interest. Sky images of this kind are made regularly, after one or just a few source frames. These sky frames are used as the flat field with the following advantages :

- As the sky is used, the intensity and spectrum of the flat field will be very close to that of the background in the source frame. Also, as the sky backgrounds can be very bright in the infrared, objects tend to be small perturbations on the background, reducing the errors due to colour difference and non-linearity.

- The time between source and sky frames will typically be short ($\sim$ minutes), due to the relatively short on-chip integration times used in the infrared. Thus the spectrum and intensity of the flat field will have varied little from that seen in the source frame. This short on-chip exposure time means that high signal to noise flat fields may be obtained easily.

- In contrast to the 1 μm wavelength cut-off of silicon, where the sky background is dominated by the monochromatic emission lines, the long wavelength cut-off of InSb is at 5.5 μm. The background there is dominated by the smooth black-body spectrum of thermal emission from the warm optics. Thus fringing is less likely to occur, as the highly structured emission lines are dominant in the 1–2.5 μm region, where the typical thickness of InSb used in a backside illuminated detector is completely opaque.

- The limited spatial coverage of current infrared arrays ($\sim 1\hat{'} \times 1\hat{'}$) means that nearby areas of sky clear of any point sources brighter than $\sim K = 16^{m}.0$ may be found fairly easily—even if not, a few randomly centred and stacked sky frames would allow the elimination of any unwanted sources with techniques similar to those of Tyson discussed above.

- As the spatial coverage is currently limited, extended sources will often fill the field-of-view, and accurate sky determination from within the source frame may be difficult or impossible. As with AC chopping, this technique provides an almost real-time monitor of the local sky brightness, as long as intensity changes are linear between two sky exposures, and the sky value for the intervening frame can be obtained by interpolation. This technique sky determination is a compromise nevertheless — the best way to determine the sky level is from within the source image itself, as intensity variations in the sky level become irrelevant as long as all the pixels integrate over the same variations.

As we have seen in section 3.9, the flat fielding of the SBRC arrays in IRCAM is extremely good, no doubt due to low levels of inherent spatial non-uniformity in the detector itself, and the high quality of the flat fields provided by the process discussed above. There are improvements yet to made to this process, such as median stacking and spatial offset frames, and the flat fielding accuracy will be enhanced further.

Integrating infrared arrays that reference the sky infrequently could potentially gain a factor of two in total integration time and root two in background limited sensitivity over chopping AC photometers. Using the flat fielding technique we have outlined above, this gain is lost, as we spend just the same fraction of the time on the sky with the array as we would with the photometer. By stacking a whole nights worth of flat fields, we could retrieve the factor of root two in sensitivity. We should be able to gain yet another factor of root two, if the flat fields could be obtained off line, during dusk or dawn. As we have discussed above however, dome flats are unlikely to be as good as sky flats, and the factor of two may be lost again. Clearly then, we need to assess the quality of the various flat fielding techniques more rigorously, in order to properly define the advantages and disadvantages of each technique. As with optical CCDs, much effort will be expended in finding better and better ways to flat field infrared array detectors.

3.11 Conclusions

In this chapter, we have developed a model of the detector physics that govern the operation of the SBRC 62 × 58 InSb + DRO array as used in IRCAM and other astronomical near infrared imaging cameras. This model has allowed us to explore the parameter space within which useful astronomical data may be obtained. We have used the model to examine the inherent non-linearity of the SBRC array in detail. Using a definition of non-linearity that applies to the the signals measured from astronomical sources *after* data reduction, we have found that the non-linearity is well behaved and below 10% in virtually all observing regimes.

We have discussed techniques that may be used to remove the non-linearity, and have developed practical implementations of a numerical technique that assumes no prior knowledge of the detector physics, and an analytical approach which is applicable to the self integrating detector only. Both techniques have been applied to model data and have been found to correct well for the non-linearity, in all but the most extreme cases, with 3% and 5% being the maximum residual non-linearity left after correction by the numerical and analytical techniques respectively. The residual non-linearity was found to be below 1% for a wide range of likely situations.

We have also developed a noise model for integrating array detectors, taking into account the propagation of noise through the astronomical data reduction process. Using this model, we have analysed some typical imaging data, and have found that the noise in the background after reduction is completely explained by read noise, and sky and dark current shot noise components. When using a sky frame taken close in time to the source frame, almost perfect flat fielding is seen to be possible, with no contribution to the final noise from residual spatial non-uniformity across the array. This excellent background limited imaging performance is consistent with the architecture of the infrared hybrid array, the nature of the sky background, and the frequency of obtaining sky frames.

To conclude : from a theoretical point of view, the SBRC array exhibits excellent characteristics for astronomical near infrared imaging. The main problem is with non-linearity, but this is generally correctable. Limited available data from detector arrays currently in operation indicate that the modelling carried out in this chapter describe the arrays well. Further effort is required to fully understand the nature of

low temperature dark current sources, that dominate below 40 K, when diffusion and generation-recombination currents have been cooled out. Finally, more data is required from actual arrays in a laboratory environment, in order to fully confirm or deny the modelling presented here, and a rigorous characterisation process should be carried out, particularly at low temperatures and low photon backgrounds, as will be encountered in future spectroscopic instrumentation.

Chapter 4

SIRCAM

4.1 Introduction

As we have seen in previous chapters, the problems associated with infrared imaging are many. The nature of the sources observed, the constraints imposed by ground-based astronomical techniques, and the relative complexity of the detector technology employed, all create a large number of possible observational circumstances. To anticipate the problems and limitations that will arise in all these situations is not a simple task. Whilst designing and implementing a first generation infrared imaging camera, we have simultaneously developed a software simulation system, known as SIRCAM (for Simulated IRCAM), to help us do so. The system has been described briefly elsewhere [95], but in this chapter we shall examine in detail the motivations and methods of, and some sample results from SIRCAM.

4.2 Motivation

The SIRCAM system was developed in order that the many possible observational configurations of an infrared imaging system could be explored, with these motivations in mind :

- To provide input during the design stages of the IRCAM system, such that the optimally efficient astronomical instrument would arise

- To estimate the likely levels of photon flux illumination arriving at the detector from both astronomical and background sources, in order that the actual system could be checked for poor throughput and background leakage

- To derive predicted figures of merit for various observational configurations in such a way as to allow the real camera system to be configured in the most efficient manner for any given application

- To model the behaviour of the detector array, in order that its operation be optimised

- To examine the calibration algorithms necessary to reduce data output from the detector array to relative photon flux units

- To provide realistic data with known input parameters, in order to test the effectiveness of data reduction and analysis techniques

In the first part of this chapter, we shall examine the way in which input astronomical photons are converted into detected infrared images by a camera. A 'recipe' is given for calculating the effects of the various components that go to make up a ground-based infrared imaging system, resulting in a photon flux falling on the detector array. This flux is detected and converted into an electronic signal by the array, this signal then being stored and examined in a computer. Throughout, examples are given that generally apply to the IRCAM 1–5 μm camera, which employs an SBRC 62 × 58 InSb + DRO array, and which is used on the 3.8 m United Kingdom Infrared Telescope, on Mauna Kea, Hawaii. However, the SIRCAM system has been implemented in a generalised and modular fashion, and should in principle be able to simulate a large range of other infrared instrumentation. It has in fact been used elsewhere to simulate the performance of single-element infrared photometers, designs for mid infrared cameras and instrumentation for future very large telescopes, and has been adapted to simulate infrared spectrometer designs based around two dimensional detector arrays [94,151].

The second part of the chapter presents some of the results output by the SIRCAM system. We give predicted background fluxes and some guide to their possible variability. In particular, various estimates of the sensitivity of IRCAM are given, across its large range of operating configurations, in order to provide a guide to observing with the camera. Also, some realistic simulated images are presented, in order to provide some

feeling as to the quality of the data that can be expected from IRCAM in its many modes. Finally, we discuss some of the problems that may arise in the low spatial resolution modes of IRCAM, particularly with regards photometric accuracy, and aliasing.

4.3 Components

Here we shall list the component parts of SIRCAM which are simulated in order to convert input astronomical photons into an electronic signal.

- Astrophysics — the simulation of astrophysically realistic numbers of photons arriving above the atmosphere for a variety of sources
- Atmosphere — the effects of attenuation, emission, and blurring processes in the atmosphere above the telescope
- Telescope — the way in which flux is collected by the optical components of a telescope, and the attenuation and emission that arises in it
- Camera — a description of the cold optical components of a cryogenic camera, including re-imaging plate scales and cooled interference filter profiles
- Detector — the way in which the detector array responds to input photons, and the sources of dark current and noise inherent in the detector

The last section — the detector — is dealt with in detail in chapter 3. The modelling of a generalised infrared focal plane array has not been attempted at this stage, due to the complexity and differing approaches of the various manufacturers involved. Thus, the only detector model dealt with in this thesis is that for the SBRC 62×58 InSb + DRO array, the array used by IRCAM. The SIRCAM system has been coded in such a way, however, that it will be possible to interface future detector models to the back-end of the other simulation modules. In this chapter, we merely use the results of chapter 3 to provide the conversion of photons arriving at the detector array into output electronic signals. These numbers, combined with the noise model developed in chapter 3, are used to derive the sensitivity figures presented later in this chapter.

4.4 Astrophysics

In the first section we shall detail the various ways in which astronomical objects are 'simulated' in SIRCAM; that is, the ways in which an astrophysically meaningful image may be created and input to the SIRCAM system. In principle, many parameters could be associated with such an image — each point on the sky could be described in terms of its spectrum, polarisation, time variability, and so forth. In order to simplify the problem, we have considered an image to be represented by the number of photo-electrons per second measured at the detector array. Next we discuss how the correct number of detected photo-electrons per second per pixel is calculated.

4.4.1 Model stellar spectrum

Although SIRCAM is capable of working at high spectral resolution, many of the simulations are performed considering only results pertinent to broad-band filter measurements. As such, only a relatively simple stellar flux model is used to generate the number of photons across the spectral bandwidth requested. An empirically derived black-body formula for the spectrum of Vega is used as detailed below.

Vega is a bright, nearby A0V star, and is used to zero-point the astronomical magnitude system by defining Vega to have a brightness of $0^m.0$ at all wavelengths. The brightnesses of all other stars are calibrated in relative terms through this definition, at least for broad-band filter measurements. Considerable effort has been expended in attempts to obtain an absolute flux calibration for Vega at all wavelengths, including at near-infrared wavelengths [111]. Vega has an infrared excess longward of $1.24\,\mu$m when measured fluxes in this range are compared to values derived from model stellar atmospheres and interferometrically measured stellar diameters. This excess is probably due to a cool dust shell [111]. By comparing the flux measured from Vega with that measured from a calibrated black-body furnace across the near infrared, Mountain *et al.* [111] have derived an empirical black-body formula that is accurate to within a few percent across the near-infrared, at least for broad-band filters. The derived black-body formula has been re-normalised into SI photon units, and is given in equation 4.1 :

$$I_\lambda = \frac{9.885002 \times 10^{-8}}{\lambda^4(\exp(1.4388 \times 10^{-6}/\lambda) - 1)} \qquad \text{photons/sec/m}^2\text{/m} \qquad (4.1)$$

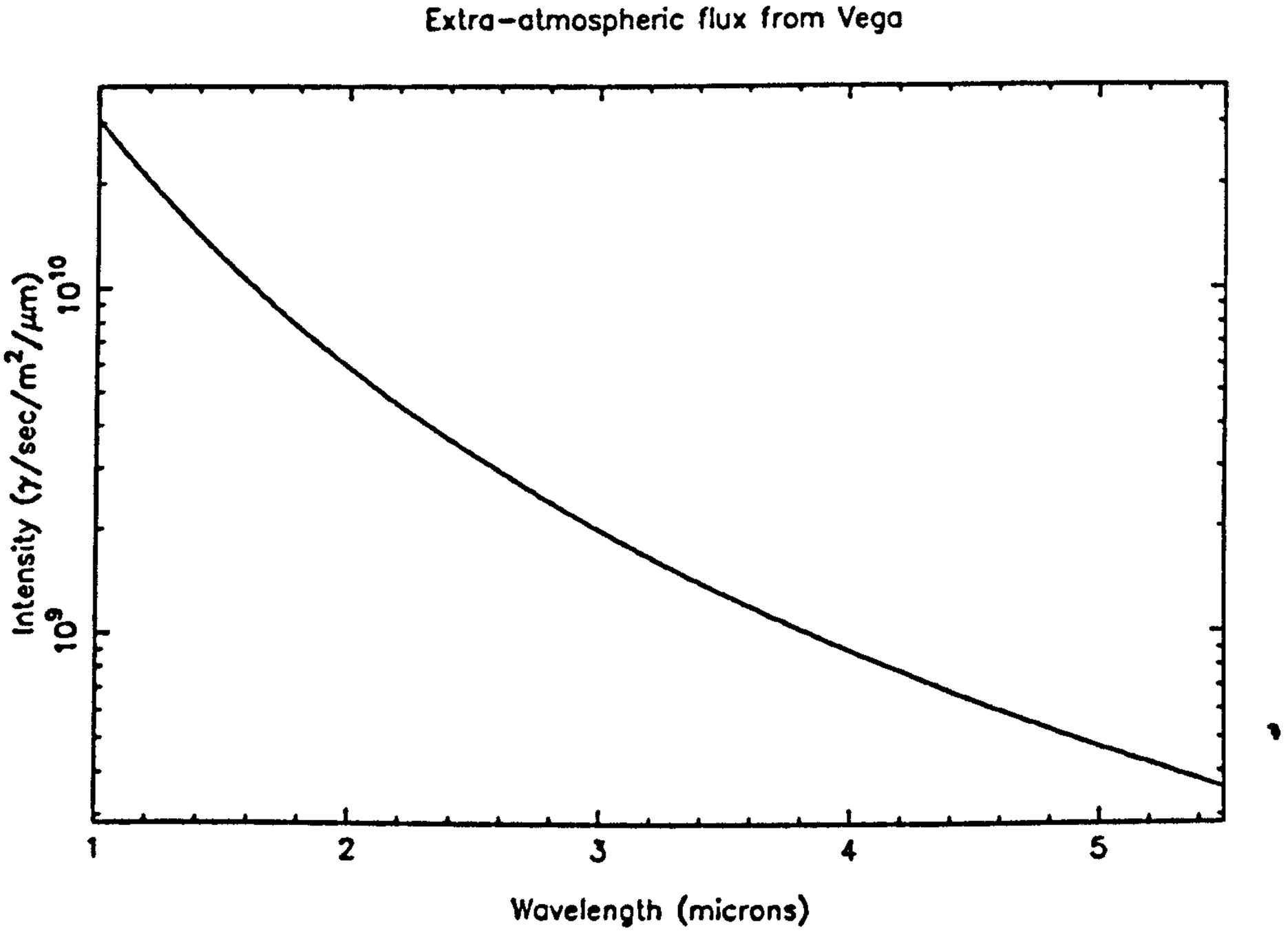


Figure 4.1: Extra-atmospheric flux from a $0^m.0$ star

Although the empirical formula given in equation 4.1 has no physical significance and will not be accurate in the presence of strong stellar lines in Vega's spectrum (e.g. Pβ at 1.282μm), it is nevertheless a useful way of calculating the extra-atmospheric flux from a zeroth magnitude star. For a star of magnitude m_{star} through a given filter, the intensity I_{star} is found in the usual way from the corresponding value for Vega through the same filter, I_{Vega} :

$$I_{star} = \frac{I_{Vega}}{10^{(m_{star}/2.5)}} \tag{4.2}$$

Figure 4.1 shows the number of photons arriving above the atmosphere per second per square metre of collecting area per micron of wavelength from a $0^m.0$ star across the near-infrared, as derived from equation 4.1.

At this stage, SIRCAM runs in a spectral mode. The model stellar spectrum is calculated for a star of the required brightness, and the resulting spectrum is then convolved with all the atmospheric and instrumental absorption, transmission, reflection, and quantum efficiency parameters discussed later in the chapter. This gives a detected spectrum; i.e. for each bin of the spectral range, the number of photo-electrons detected per second is known. The final step is to convolve this spectrum with the profile of which ever filter has been chosen. The resulting spectrum is then binned to give the

total number of photo-electrons per second detected for the given star through the given filter. SIRCAM now switches to spatial mode, where the given detected flux is spread throughout a 2-d image as discussed below.

This simple empirical model has some problems simulating the flux arriving from cool or heavily embedded sources. The problem arises due to the fact that magnitudes of sources are generally referred to a single monochromatic wavelength, and the definition of this effective wavelength relies on the assumption that the spectral dependence of the source across a wide bandpass filter is substantially similar to that of the standard star across the same bandpass, and that the atmospheric transmission does not change dramatically across the bandpass. The spectrum of most stars can be approximated by a single black-body. A black-body at a temperature of 3000 K will peak at about 1 μm, and will drop according to the Rayleigh-Jeans approximation (λ^{-4}) into the infrared. Thus, all stars with an effective black-body temperature $\geq$ 3000 K will decline in the infrared with the same spectral dependence — virtually all stars lie above this temperature. For these stars, a linear scaling of the flux from Vega according to equations 4.1 and 4.2 will be accurate for any given infrared wavelength. However, for sources with black-body temperatures below 3000 K (e.g. solar system objects) or with spectral dependencies other than that of a black-body (e.g. a source emitting with a power law spectrum, or a source deeply embedded in dust), this scaling may break down. In order to correctly predict the flux from an object with a significantly different spectral dependence from that of Vega across a wide bandpass, a colour correction term is required. These colour corrections are treated in more detail elsewhere [144,59,63], and it is seen that the colour corrections to be applied may be as little as a few percent for cool stars through the standard filters, ranging up to factors of two or three for very cool objects. In the context of this work, however, we shall ignore these corrections. All we are attempting to do at this point is simulate imaging systems, and thus the basic figure of interest is the total number of photons detected from a given source through a given filter. Using the Vega model for reddened or cool sources may result in slight errors in this number, but we really only require it to be correct to the right order, to say $\pm$25%. From the model spectrum, we know how many photons per second we input to the system, and as long as we use a consistent model throughout, we should be able to work back from the number of detected photo-electrons to the input photon flux. Work in this area would be necessary in order to properly simulate spectroscopic applications.

4.4.2 Point sources

The simplest image to simulate is a field of point sources. In order to be able to simulate images of the same star field at various spatial resolutions and through any of the broad band filters, we use simple catalogues. These catalogues hold a defined number of stars, each with a position on the sky in some arbitrary units, and each with a list of broad band magnitudes. When we come to generate an image, the pixel scale, broad band filter, and position on the sky are input, and from the catalogue, the brightness of each star and its position relative to the detector array are calculated. From the model stellar spectrum above, and the wavelength dependent optical transmission and reflection calculated below, the number of detected photo-electrons per second is calculated for each star.

Having calculated the number of photo-electrons per second detected by the detector array that represents a star of a given magnitude, we now have to spread this flux across the array. This is done according to a seeing profile. The form of the seeing profile used by SIRCAM is discussed below. After having input the flux from each star, a constant background flux is added, as calculated from the SIRCAM background predictor. An example simulated star field is shown towards the end of this chapter.

4.4.3 Extended emission images

Extended emission is often seen in the infrared. The large area coverage, large number of pixels, and simultaneous integration qualities attached to infrared arrays make them ideal for imaging extended emission at high spatial resolution. We need to simulate such extended emission at various intensity levels in order to provide some feel for the sensitivity of a camera operating at different image scales, noting also that the time and intensity sensitivity extrapolations that apply to point sources are different for extended emission.

In creating simulated extended source images, the following reasons determined the image size and spatial structure scale. At 2.4″/pixel, the SBRC array of 62×58 detectors subtends a field-of-view of $\sim 150'' \times 150''$, neglecting vignetting. Thus, we require our simulated image to be at least that big. The highest spatial resolution mode of IRCAM is 0.6″/pixel. At this scale, we roughly fully sample the image blur due to seeing. Thus, details in the original (extra-atmospheric) image are required at or below the

seeing level. Taking both requirements together, an image of size 512 × 512 pixels with structure scales as small as one or two pixels, roughly satisfies the demand. Several such images were created by overlaying random polynomials, smoothing to get rid of edges, and then running a deconvolution algorithm to sharpen up the (non-existent) original details. Their actual structure is astrophysically meaningless, but do have structure at the required spatial scale, and have a typical dynamic range of surface brightnesses.

To convert the standard image into an input SIRCAM image, the following steps are followed. Firstly, depending on the chosen image plate scale, a sub-section of the image is extracted, corresponding to the field-of-view of the detector array. Then, a brightness range of the input source at the required wavelength is specified. Using the Vega model spectrum convolved with the optical components and detector quantum efficiency discussed below, we calculate the number of photo-electrons that would be detected per second per pixel from a given contour in the model surface brightness image, and the image is multiplied by the relevant scalar to normalise its intensities to the correct extra-atmospheric levels. The correct background flux level for the given wavelength and pixel scale is also calculated, and added as a uniform constant to the array. Finally, the image is smoothed with a seeing profile. Examples of simulated extended emission images are shown towards the end of this chapter.

4.5 Atmosphere

In this section, we shall discuss the rôle played by the atmosphere whilst making astronomical observations in the infrared, and the ways in which the various components are simulated in SIRCAM. The three basic aspects we shall look at are :

- the 'seeing', the effective blurring applied to the incoming astronomical photons by turbulence in the atmosphere

- the attenuation of astronomical light by absorption in the atmosphere

- the infrared radiation emitted by the atmosphere, in thermal equilibrium and otherwise

4.5.1 Seeing

Atmospheric turbulence degrades the spatial resolution that can be obtained with an imaging system, well below the level set by the diffraction limit of all but the very smallest telescopes. The diffraction limit ($1.2\lambda/D$) for a 4 m optical telescope operating at 0.5 μm is $\sim 0.03''$, whereas the typical full-width at half-maximum (FWHM) of an atmospheric blurred point source is $\sim 1''$. This atmospheric blurring is known as the 'seeing', and is due to fluctuations in the refractive index of the air in turbulent elements, where the turbulence is described by a Kolmogorov power spectrum — full reviews of seeing theory are given elsewhere [168,30]. There are three parameters of importance that we shall discuss here :

- the typical seeing for Mauna Kea
- the wavelength dependence of the seeing
- the functional form of the seeing profile

4.5.1.1 Optical seeing on Mauna Kea

The median seeing at optical wavelengths on Mauna Kea is about 1–1.5$''$ FWHM [152]. However, for a substantial fraction of the time ($\sim$ 30%), the seeing is less than 1$''$ FWHM, with values lower than 0.5$''$ reported as often as 15% of the time at the CFHT [1, page 5-1]. High winds worsen the seeing, as do poor thermal control in the dome, guiding errors, and other image motions; it has been suggested that the 'intrinsic' site seeing may be as good as 0.4$''$ FWHM on average [1, page 5-2].

4.5.1.2 Wavelength dependence

An important figure of merit associated with the seeing is Fried's parameter r_0, which is roughly numerically equivalent to the diameter of a telescope which would have an Airy disk of the same size as the disk of a star blurred by seeing. Thus, as the seeing gets better, r_0 gets bigger — for example, for 1$''$seeing, $r_0 \simeq 5$ cm, and for 0.5$''$seeing, $r_0 \simeq 10$ cm. The r_0 parameter is weakly wavelength dependent, with $r_0 \propto \lambda^{6/5}$ [30]. As the RMS angular blur depends on λ/r_0, the seeing exhibits a $\lambda^{-1/5}$ dependence.

Thus, for seeing of 1″ FWHM at 0.5 μm, the seeing at 2 μm will have improved slightly to $\sim$ 0.75″ FWHM. The diffraction limit for a 4 m telescope at 2 μm is 0.12″ — the gain in seeing in the infrared is really only appreciable for telescopes larger than 1 m at wavelengths longer than 2 μm, as the diffraction limit trades off against the seeing improvement. In order to take full advantage of these gains with high resolution imaging instrumentation, however, infrared telescopes need to be figured to the same high accuracy as optical telescopes [25].

4.5.1.3 Point source profile

In order to simulate the effect of the atmosphere on point sources and high spatial resolution extended emission, we need to know the approximate functional form of the seeing. A Gaussian profile is frequently assumed, but it is known that a much better fit is obtained through the use of a Lorentzian of the form [33] :

$$I(x,y) = \frac{1}{1+\alpha^{\beta}} \tag{4.3}$$

where $\alpha = \sqrt{\left(\frac{x}{r_x}\right)^2 + \left(\frac{y}{r_y}\right)^2}$

$\beta = P(1+\gamma)$

$\gamma = \sqrt{\left(\frac{x}{p_x}\right)^2 + \left(\frac{y}{p_y}\right)^2}$

x is the distance in the x direction from the profile centre

y is the distance in the y direction from the profile centre

r_x, r_y are fitted radii in the x, y directions

p_x, p_y are fitted scale factors in the x, y directions

A Lorentzian profile of the form shown in equation 4.3 fits stellar profiles as observed with linear solid-state detectors much better than a Gaussian, accounting well for the observed extended wings [33]. The profile is also known to be a close approximation to the theoretical profile given by a telescope with a central obscuration looking through a turbulent atmosphere characterised by the Kolmogorov power spectrum [118]. By fitting the profile given in equation 4.3 to stellar images as measured with an optical CCD detector, with seeing between 1 and 4.5″ FWHM, Diego [33] has derived values for the fitted parameters r_x, r_y, p_x, p_y, and P as a function of the seeing FWHM — these

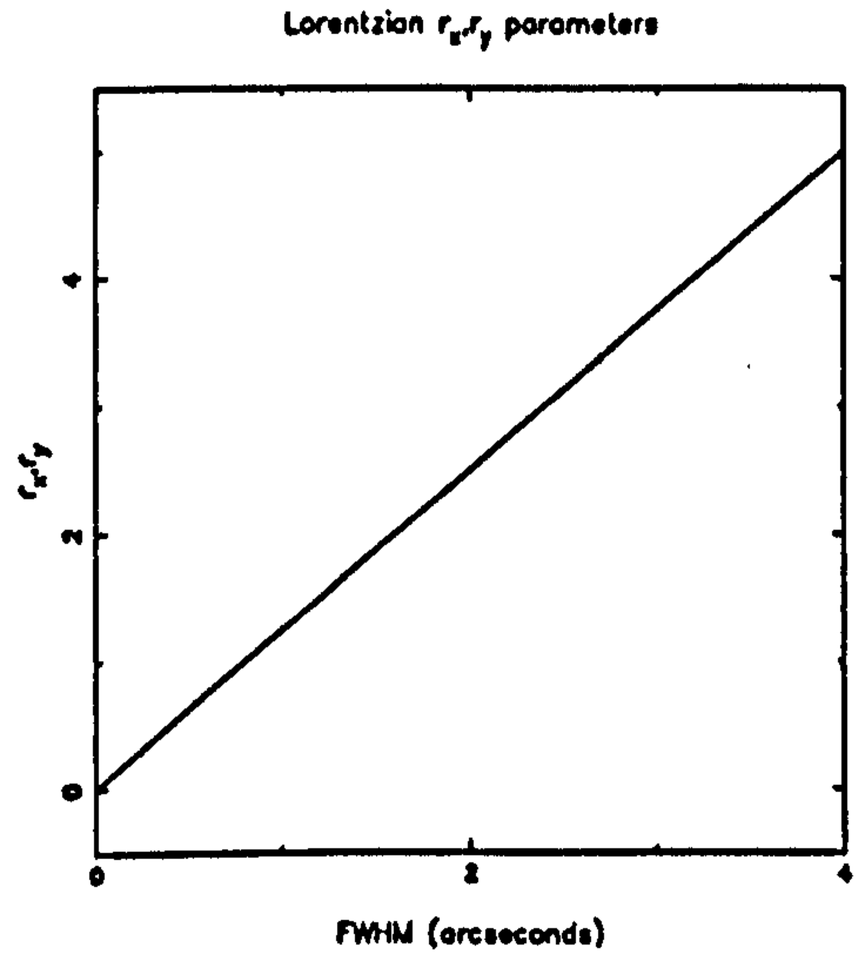


Figure 4.2: r_x, r_y versus FWHM

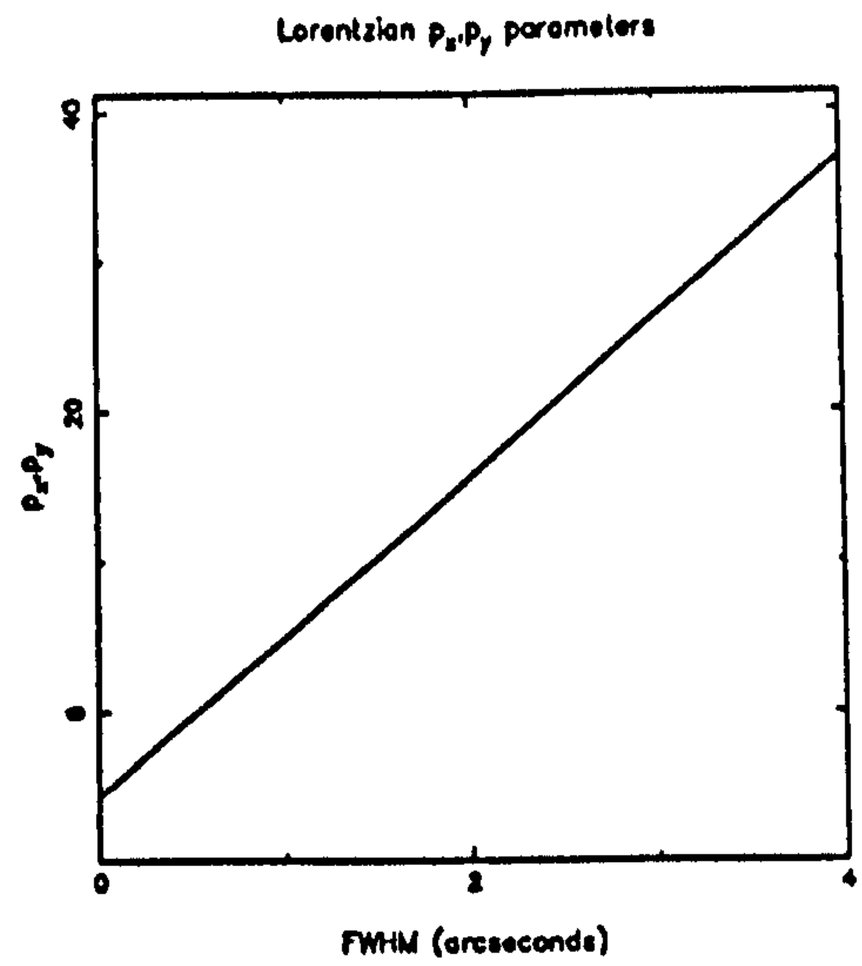


Figure 4.3: p_x, p_y versus FWHM

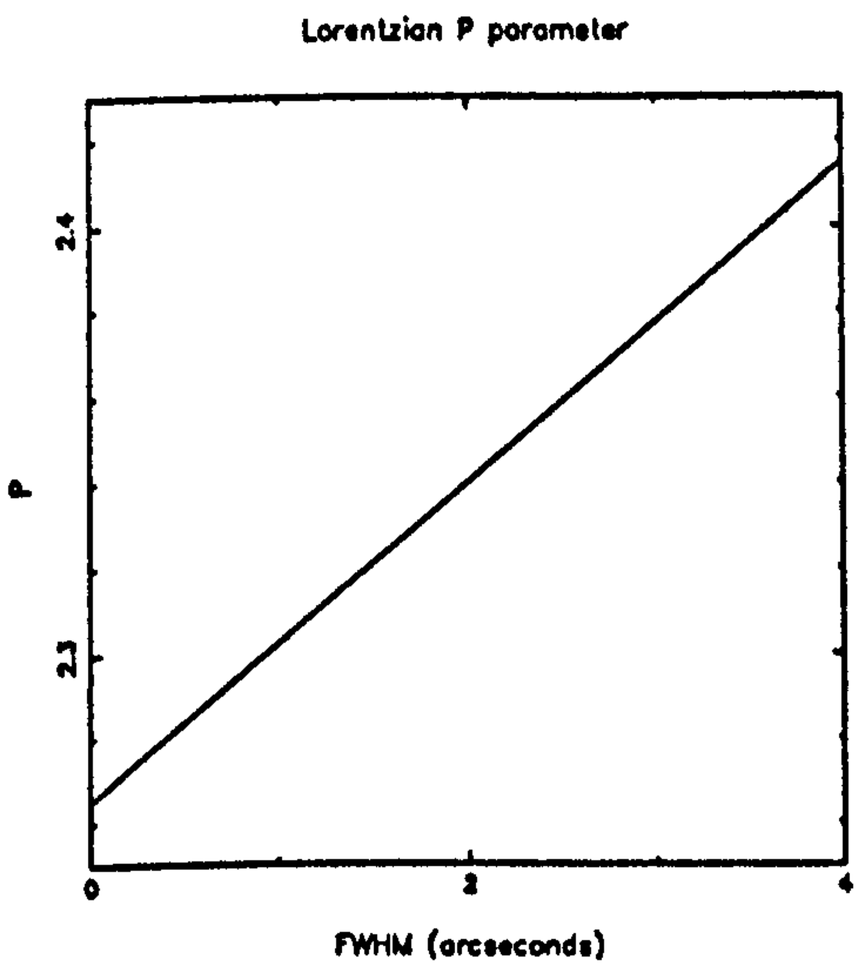


Figure 4.4: P versus FWHM

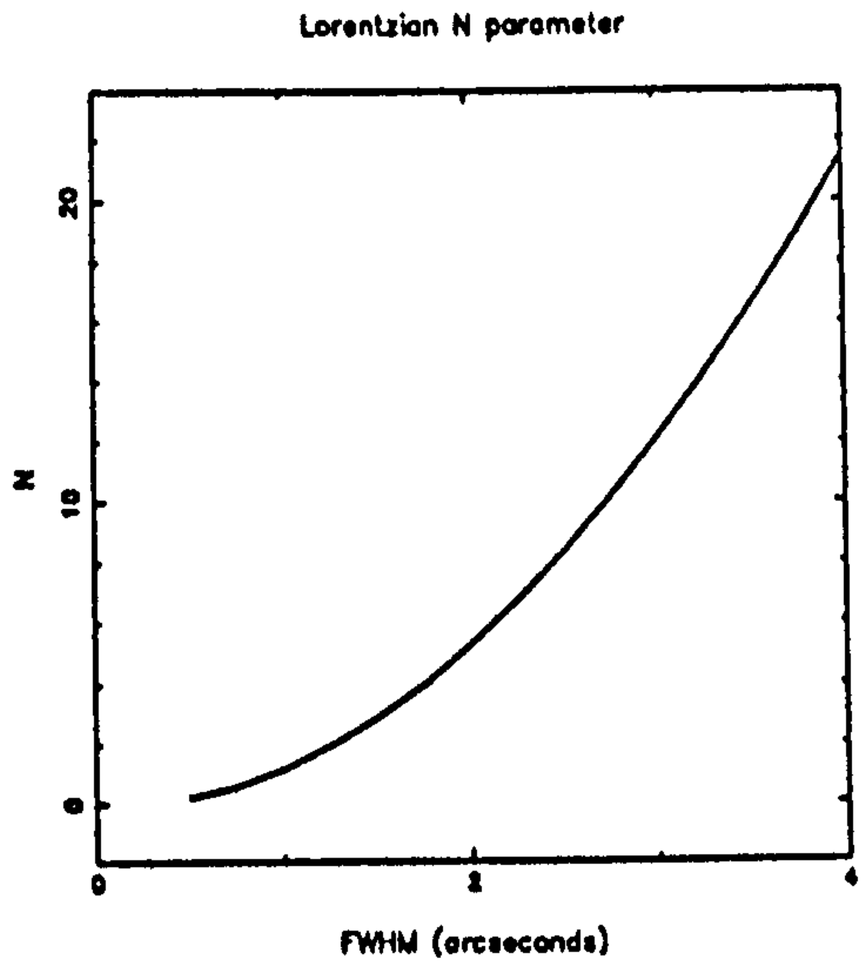


Figure 4.5: N versus FWHM

are shown in figures 4.2 to 4.5. These are the values used by SIRCAM when simulating a point spread function with a given FWHM. It should be noted that Diego omitted one important point from his paper : the parameters shown in the figures are in units of $0.33\,\widehat{''}$, the size of the CCD pixel used to derive them, and the user must be careful to convert the x and y coordinates to these units before generating the Lorentzian profile (Diego, private communication).

Two problems arise when simulating a stellar profile. Firstly, the Lorentzian profile of equation 4.3 is not easily normalised; that is, it is not simple to integrate the function between $+\infty$ and $-\infty$ in both x and y to determine the normalisation factor N that sets the volume integral to one. We need such a normalisation factor in order to be able to spread a known amount of flux over the profile. Given the changing nature of the profile with the FWHM, the only way to obtain such a normalisation factor is

numerically. By generating profiles with the form of equation 4.3 using the parameters of figures 4.2–4.4, truncating the profiles at a radial distance of 10× FWHM where $I' \simeq 10^{-6} I_0$, and then summing the flux under the profile, the normalisation factor N was empirically determined as a function of the seeing FWHM, and is shown in figure 4.5. In order to allow simple calculation of the normalisation parameter N as a function of the seeing FWHM, a polynomial was fitted to the data shown in figure 4.5. A cubic fit was used, with a mean deviation of the normalisation parameter determined by the fit being about 0.3% from the actual normalisation parameter :

$$N = -2.72020 \times 10^{-2} F^3 + 1.51278 F^2 - 0.233191 F - 7.02060 \times 10^{-2} \qquad (4.4)$$

where F is the full width at half maximum (FWHM) of the seeing profile. Thus, for the specified seeing FWHM, the volume normalisation parameter is evaluated, and used to ensure that the sum of the flux under the profile is normalised to one.

The second problem is due to undersampling. The IRCAM plate scales are such that the seeing disk is only marginally fully sampled (0.6″/pixel) or heavily undersampled (1.2,2.4″/pixel). By evaluating the intensity of the Lorentzian at the centre of each pixel and assuming that the mean intensity for the whole pixel is the same, leads to serious errors as the profile changes significantly and non-linearly across the pixel. Thus the concept of 'virtual pixels' was introduced. That is, each real (physical) pixel is sub-divided into an array of virtual pixels, where the fineness of the sub-division is determined by the seeing-to-pixel scale ratio. The Lorentzian profile calculation is performed for each virtual pixel in turn, for each of which the change in the profile is small and can be assumed to vary linearly, before binning up all the virtual pixel values to derive the flux falling on the real pixel. It has been empirically determined that a virtual pixel size 1/10th the seeing FWHM gives very accurate results — this is equivalent to five times oversampling, and is equivalent for a Gaussian profile to four virtual pixels per σ. For 1″ FWHM seeing with a plate scale of 0.6″/pixel, each real pixel is divided into a 6 × 6 array of virtual pixels, whilst at 2.4″/pixel, a 24 × 24 sub-array would be used.

The concept of virtual pixels is useful elsewhere, particularly when considering the profile of a pixel, where cross-talk, dead-space, and intra-pixel variations are relevant. For example, in sections where we refer to fill factor, we simulate a dead space around the outside edge of the detector using virtual pixels that are defined as having zero

quantum efficiency.

4.5.1.4 Seeing used by SIRCAM

The SIRCAM model uses a Lorentzian profile as described above. Taking into account the slight gain with wavelength, and the fact that the UKIRT is unlike the CFHT and not an optimum optical telescope, values of 1–1.5″ FWHM are typically used in SIRCAM simulations.

4.5.2 Atmospheric transmission

Astronomical photons will be transmitted by the atmosphere to varying degrees, dependent on the wavelength of the photons. The optical region from 0.35–0.8 μm is almost completely transparent, but the atmospheric transmission spectrum becomes complex as we enter the infrared longward of 1 μm, with several windows and opaque regions defined by various molecular species, notably water vapour and carbon dioxide. In order to calculate the fraction of the extra-atmospheric flux that is transmitted for any given wavelength and thus available for detection by the telescope and instrument, we need to know the spectral dependence of the atmospheric absorption. Also, the non-thermal processes which give rise to the so-called 'nightglow' (see section 4.5.4) occur in the mesopause, well above the lower atmosphere or troposphere where the great majority of the absorption takes place. Thus we also need to know the atmospheric transmission spectrum in order to calculate the night sky brightness between 1 and 2.5 μm. Finally, the atmosphere radiates as a black-body, with higher emissivity in the more opaque regions of the spectrum. As shown in section 4.5.3, we can use the transmission spectrum of the atmosphere to calculate the black-body flux emitted, which becomes important longward of 2.5 μm.

Rather than using a measured atmospheric transmission spectrum, we have chosen to use model spectra, as generated by the IRTRANS model, fully described by the authors, Traub and Stier [145], elsewhere [110], and summarised here.

IRTRANS uses the Air Force Cambridge Research Laboratories (AFCRL) atmospheric absorption line parameter data base, which holds information on over 109,000

known transitions of O_2, O_3, H_2O, CO, CO_2, N_2O, and CH_4, which have wavelengths between 0.76 μm and 3.26 mm. The data base lists the wavenumber, line strength, pressure broadening coefficient, and lower state energy level—the IRTRANS model then takes these data along with input column abundances, effective temperatures and pressures of each molecular species, and calculates the atmospheric transmission spectrum. To do this, the model assumes a single isothermal layer (Curtis-Godson) approximation, in which the well-mixed species (O_2, CO, CO_2, N_2O, CH_4) are assumed to have an effective pressure one half that of the base pressure, and for H_2O and O_3, the effective pressure is assumed to be the atmospheric pressure at which the column abundance of each species has dropped to half its base value. The latter two species are not well mixed, with water vapour lying at the bottom of the atmosphere, and ozone occurring high in the atmosphere only. This approximation is known to be very accurate for both weak and strong lines, with intermediate lines showing errors of up to 5% in the line core. For well-mixed species, the mean error for well-mixed species is well under 1%, when averaged across a line profile. It would be useful, however, to replace this Curtis-Godson approximation with a multi-layered model when deriving the atmospheric emission, as the atmosphere is not truly isothermal (see section 4.5.3).

The absorption line profile used in IRTRANS is as follows. The simple case assumed in most cases is that of a symmetric pressure broadened Lorentzian of the type :

$$F_L(\sigma) = \frac{\alpha/\pi}{\alpha^2 + (\sigma - \sigma_0)^2} \tag{4.5}$$

where σ is the wavenumber at the given point in the line

σ_0 is the wavenumber of the line centre

α is the pressure broadened width, dependent on the pressure and temperature of the atmosphere

However, the simple Lorentzian is probably in error in the far wings of particularly strong lines, and the Lorentzian is only a first approximation. Also, thermal velocities of the molecules will result in Doppler broadening in the cores of lines, particularly those in the near-infrared. In this case, a Voigt profile is used to give the Doppler broadening :

$$F_D(\sigma) = \beta^{-1}\pi^{-1/2}\exp[-(\sigma - \sigma_0)/\beta]^2 \tag{4.6}$$

where $\beta = 4.30 \times 10^{-7}\sigma_0(T/M)^{1/2}$

Molecule	Column abundance (cm^2)	Pressure (mb)	Temperature (K)
O_2	2.66×10^{24}	300.0	228.0
O_3	7.28×10^{18}	36.4	219.0
H_2O	4.02×10^{21}	506.0	252.9
CO	8.88×10^{17}	300.0	228.0
CO_2	4.13×10^{21}	300.0	228.0
N_2O	3.18×10^{18}	300.0	228.0
CH_4	1.91×10^{19}	300.0	228.0

Table 4.1: IRTRANS parameters used for Mauna Kea—4.2 km 1.2 mm H_2O

M is the molecular weight

T is the temperature

The effective line profile used by IRTRANS is a convolution of the Lorentzian and Voigt profiles. The Voigt profile reduces to the Lorentzian profile under certain conditions, and the Lorentzian is used in such cases as it is simpler to calculate [145].

The main input variables to the IRTRANS model are the atmospheric parameters; the column abundance, effective temperature, and effective pressure for the molecular components. The main source of variation in these parameters is the altitude of the observatory site, as this determines the amount of atmosphere above the telescope. These atmospheric parameters have been determined for Mauna Kea [145,110], and are listed in table 4.1, with a typical Mauna Kea water vapour value used. Note that the effective temperatures and pressures used for the water vapour and ozone are different to those used for the other components—neither are well-mixed in the atmosphere, with water vapour lying close to the surface, and ozone being a high altitude component.

When calculating atmospheric transmission spectra for SIRCAM, all the parameters listed in table 4.1, except the amount of water vapour, were taken as fixed. In order to use SIRCAM with different atmospheric parameters, at another observatory site for example, the IRTRANS model would have to be re-run with the relevant parameters, and the output transmission spectrum substituted accordingly.

The two atmospheric variables accessible to the user of SIRCAM are the amount of water vapour and the airmass. Absorption in the atmosphere due to water vapour

is often the limiting factor in making millimetre, sub-millimetre, mid- and far-infrared observations. Less critical in the near-infrared, water vapour absorption nevertheless has an important influence on the L′ and M filters. Water vapour has a scale height of approximately 2 km — Mauna Kea is at an altitude of 4.2 km, and thus the atmospheric water vapour content there should be about 10% of the sea-level value. The median value for the precipitable water content of the atmosphere above Mauna Kea is 1.2 mm, never much above 4 mm, and for a substantial fraction of the time (25%) at or below 0.5 mm [22]. In SIRCAM, a water vapour of 1.2 mm is typically used, with any of the values 0.3,0.6,1.2,2.4, 3.6,4.8 mm available in the stored atmospheric data-base.

The second variable is the airmass, defined as the secant of the zenith angle. The path length of the atmosphere through which astronomical photons must pass increases proportionally with the airmass, as the telescope looks at greater and greater zenith angles, and therefore the atmospheric column abundances rise in proportion to the airmass. Often the assumption is made that the transmission of the atmosphere T for a given airmass A can be found from the zenith transmission T_0 by $T = T_0{}^A$. This is approximately true as long as the transmission remains unsaturated, but various atmospheric absorption lines are saturated at the zenith, or become so as the airmass increases, and this assumption breaks down. In order to provide a better airmass dependence, IRTRANS spectra were generated for a range of airmasses (1.0, 1.25, 1.5, 1.75, 2.0), between the zenith and 60° over.

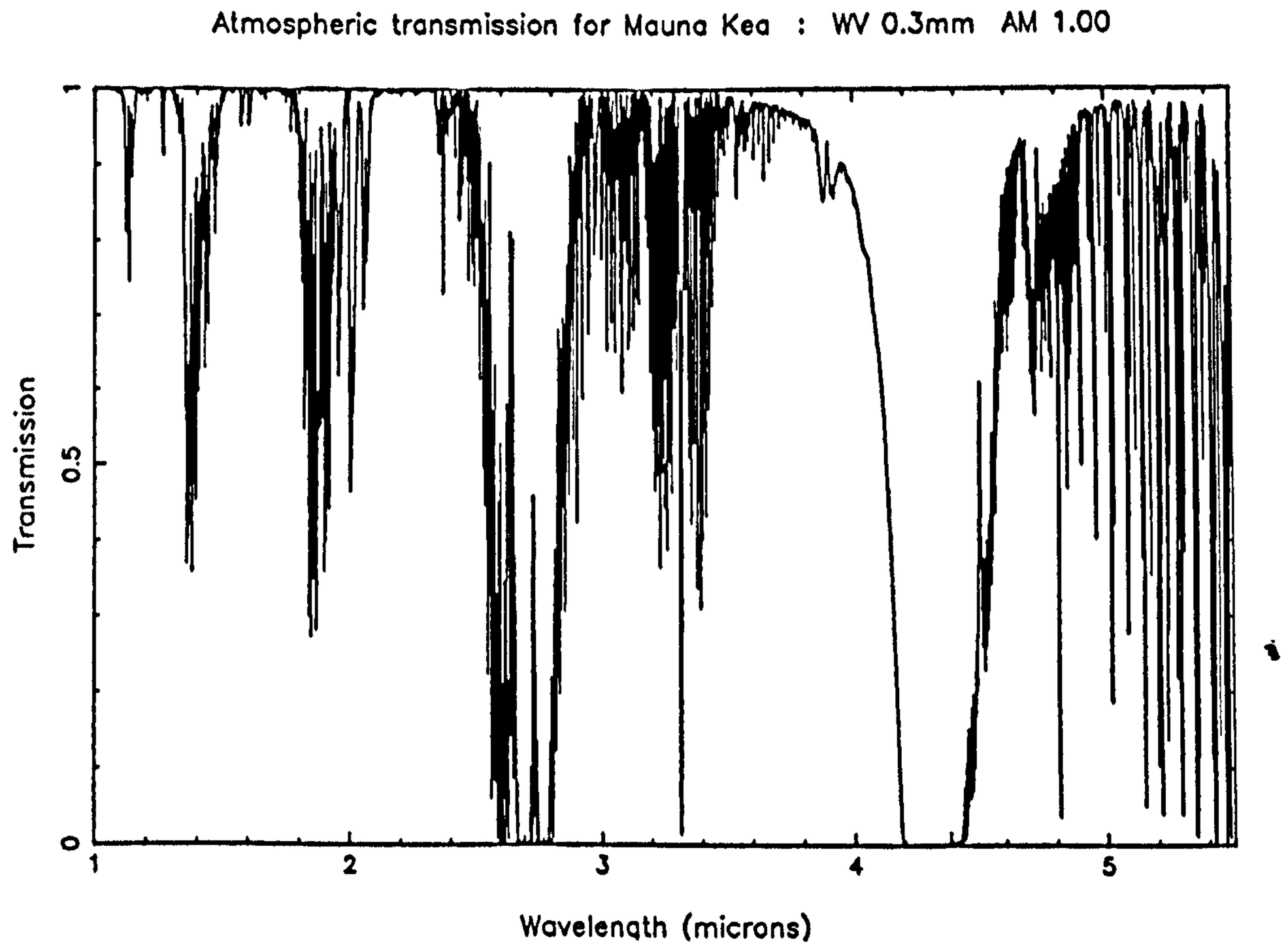
Thus, a grid of 30 1–5.5 μm atmospheric transmission spectra were generated for SIRCAM, at a spectral resolution of 10^{-3} μm. Any combination of water vapour (0.3–4.9 mm) and airmass (1.0–2.0) can be chosen by the user, and if the values are coincident with an actual calculated spectrum, that spectrum is used, otherwise the required spectrum is interpolated from the actual spectra with parameters surrounding those chosen. As the transmission is not necessarily a linear function of the water vapour and airmass, a bi-cubic spline interpolation algorithm is used (NAG routine E01ACF). Although time consuming, the interpolated spectra are high accurate. Comparing interpolated spectra with actual IRTRANS output, the interpolation with airmass is shown to be correct with a peak error of 0.5% and a mean error of 0.1%, and the interpolation with water vapour shows a mean error of about 0.5% in transmission. The errors are largest in the less transmissive regions of the spectrum, and therefore are not too important.

Figure 4.6 shows the two extreme spectra used by SIRCAM, that is the (0.3 mm water vapour, airmass 1.0) and (4.8 mm water vapour, airmass 2.0) cases, in order to demonstrate the range in atmospheric transparency that can reasonably be expected for Mauna Kea.

Note that aerosol and molecular scattering have not been included in the IRTRANS model. The effects of these components may be important at the shorter wavelengths, but are not discussed further here. More information may be found elsewhere [110,96].

4.5.3 Atmospheric thermal emission

Just as the lower atmosphere absorbs astronomical photons, it also emits black-body photons with an emissivity depending on the atmospheric opacity for any given wavelength. A rigorous thermal emission model of the atmosphere would account for the differing temperatures of the various molecular and atomic components in the various atmospheric layers, but a good approximation for the emissivity can be obtained by applying Kirchoff's Law to the IRTRANS spectra calculated with the single isothermal layer Curtis-Godson approximation, and taking one minus the transmission for each wavelength [145]. Whilst this models the emissivity well, it is not so easy to decide on one single effective black-body temperature. Above Mauna Kea, the majority of the atmospheric components have effective temperatures of about 228 K, with the exceptions of ozone (213 K) and water vapour (253 K), as shown in table 4.1. The water vapour is the main problem, as it is responsible for much of the absorption in the 3–5 μm region, where the thermal emission of the atmosphere rises steeply. Thus, in order to maximise the likely atmospheric thermal emission, the black-body emissivity is obtained as discussed above, and the effective black-body temperature is taken as the temperature of the water vapour. Errors in these assumptions are not likely to be of great importance in the near-infrared, as shortward of the room temperature black-body peak at around 13 μm, the thermal background is dominated by the telescope and warm optics, some 20 K warmer than the effective water vapour temperature. Further discussion of this complex issue may be found elsewhere [96, page 40].



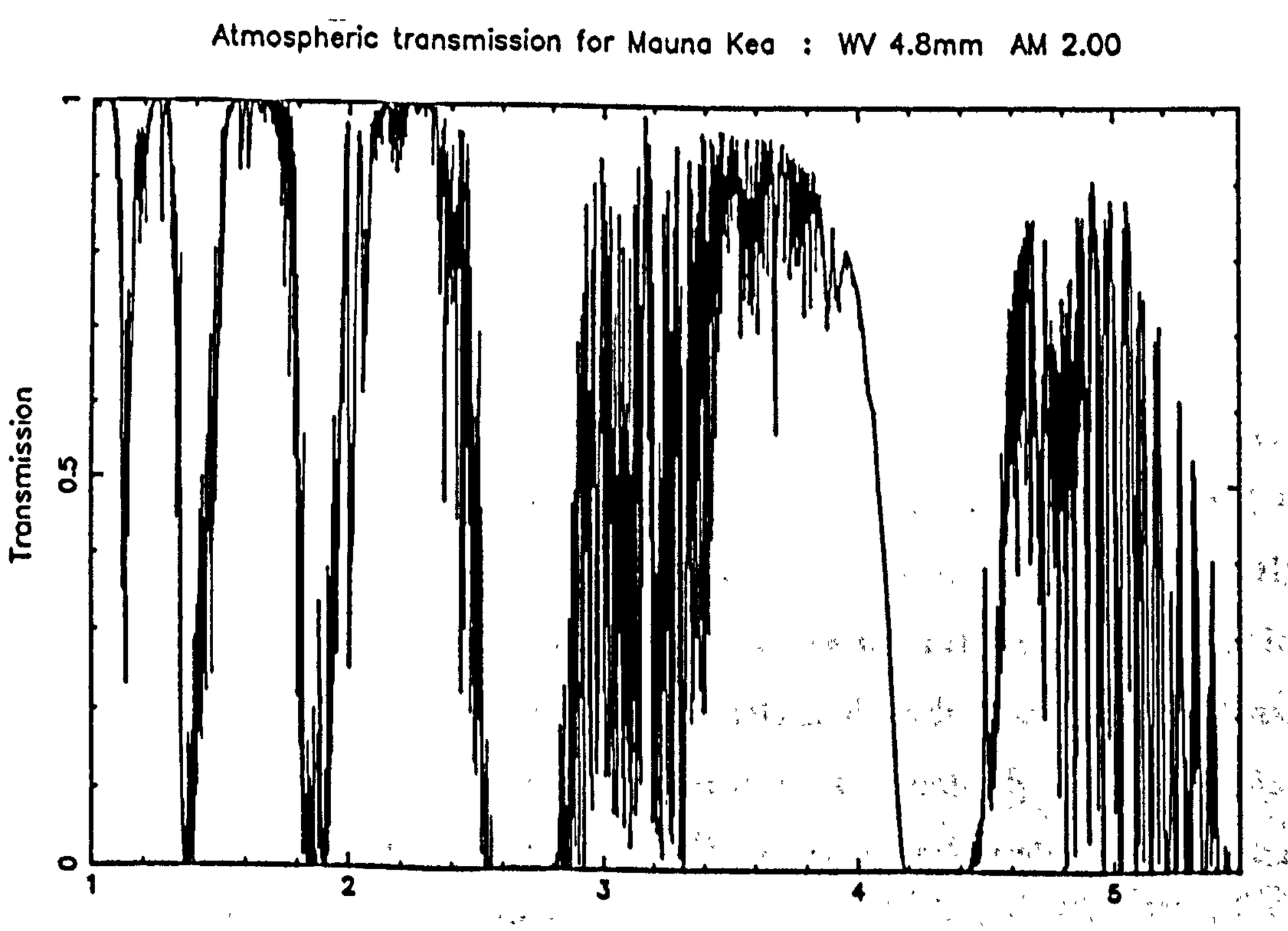


Figure 4.6: Two example atmospheric transmission spectra for Mauna Kea

4.5.4 Non-thermal emissions

There are two major sources of non-thermal radiation that dominate the near-infrared night sky from 1–2.5 μm. The first is the polar aurora, due almost exclusively in the near-infrared to emission lines from the N_2 molecule, excited by highly energetic solar wind particles hitting the upper atmosphere [148]. Even though the aurora can be very bright, it is largely confined to polar regions, reaching peak intensity between 20–25° from the Earth's magnetic poles—at the latitude of Mauna Kea (19.5° N), the aurora is almost entirely negligible, occurring over 10^3 times less frequently than in polar regions [161, page 3-79].

The dominant source of non-thermal radiation from the near-infrared sky above Mauna Kea is the so-called 'airglow', which has three main components at night, when it is known as the 'nightglow' :

- OH* vibration-rotation bands
- $O_2(a^1\Delta_g - X^3\Sigma_g^-)$ IR atmospheric bands
- the near-infrared nightglow continuum

4.5.5 The hydroxyl molecule OH*

Emission lines of the vibrational-rotational spectrum of the excited hydroxyl molecule, OH*, were first identified astronomically in the optical red by Meinel [101], and are often referred to as the Meinel bands [102]. The bands are formed as excited OH* molecules in the vibrational levels v = 1–9 relax to the ground state, OH. The OH* is formed in a thin layer of atmosphere at the boundary between the mesosphere and lower ionosphere (known as the mesopause) at about 90 km altitude from reactions between various atmospheric components that have been photochemically dissociated during the day by short wavelength solar photons. There are several production mechanisms involved [88], the major route being :

$$H + O_3 \longrightarrow OH^* + O_2 \qquad (4.7)$$
$$v' \leq 9$$

The production and subsequent quenching mechanisms for the OH^* molecule involves reactions between O, O_2, O_3, H, H_2, HO_2, and N_2, and are temperature dependent, as are final population level distribution in the excited molecule. Thus predicting the emission intensities of the various branches and lines is complex, and the theoretical spectrum is not yet well known [88,135]. However, it is felt that the smaller Δv transitions, which give rise to the infrared lines, are well modelled, and that it is the optical spectrum, arising from the the larger Δv transitions, where theoretical problems still exist (Llewellyn, private communication). In order to provide some estimate of the likely emission from this molecule in our modelling, we have adopted the following parameters as the best guess :

- the wavelengths of the bandheads (Q-branch) and R_1, R_2, P_1, and P_2 series of lines for each band that radiates in the near-infrared (the v'–v'' = 9–5 branch of the $\Delta v = 4$ group, and all the $\Delta v = 3, 2, 1$ branches) as listed by Chamberlain [28, pages 368–375]

- the relative intensities of the lines and branches within each band as calculated by Chamberlain for a mean rotational temperature of 225 K [28, pages 556–557]

- the total absolute intensities for each branch as calculated by Llewellyn, Long, and Solheim [88]

Note that the intensities of atmospheric radiation phenomena such as the airglow are usually given in Rayleighs (R), a unit describing photon emission from a volume of atmosphere in the following manner [123, page 28] :

> "If the surface brightness B is measured in units of 10^6 quanta cm^{-2} s^{-1} sr^{-1}, then in Rayleighs, R, the surface brightness is $4\pi B$."

The factor of 4π conceptually arises when a detector a swept around the surface of a sphere containing the radiating volume of atmosphere. Note that the Rayleigh involved here is Lord Rayleigh IV, who studied airglow, rather than his better known father, Lord Rayleigh III, who studied scattering.

In order to make the units more applicable to SIRCAM, the OH^* emission intensities given above have been converted from Rayleighs into photons per second per square

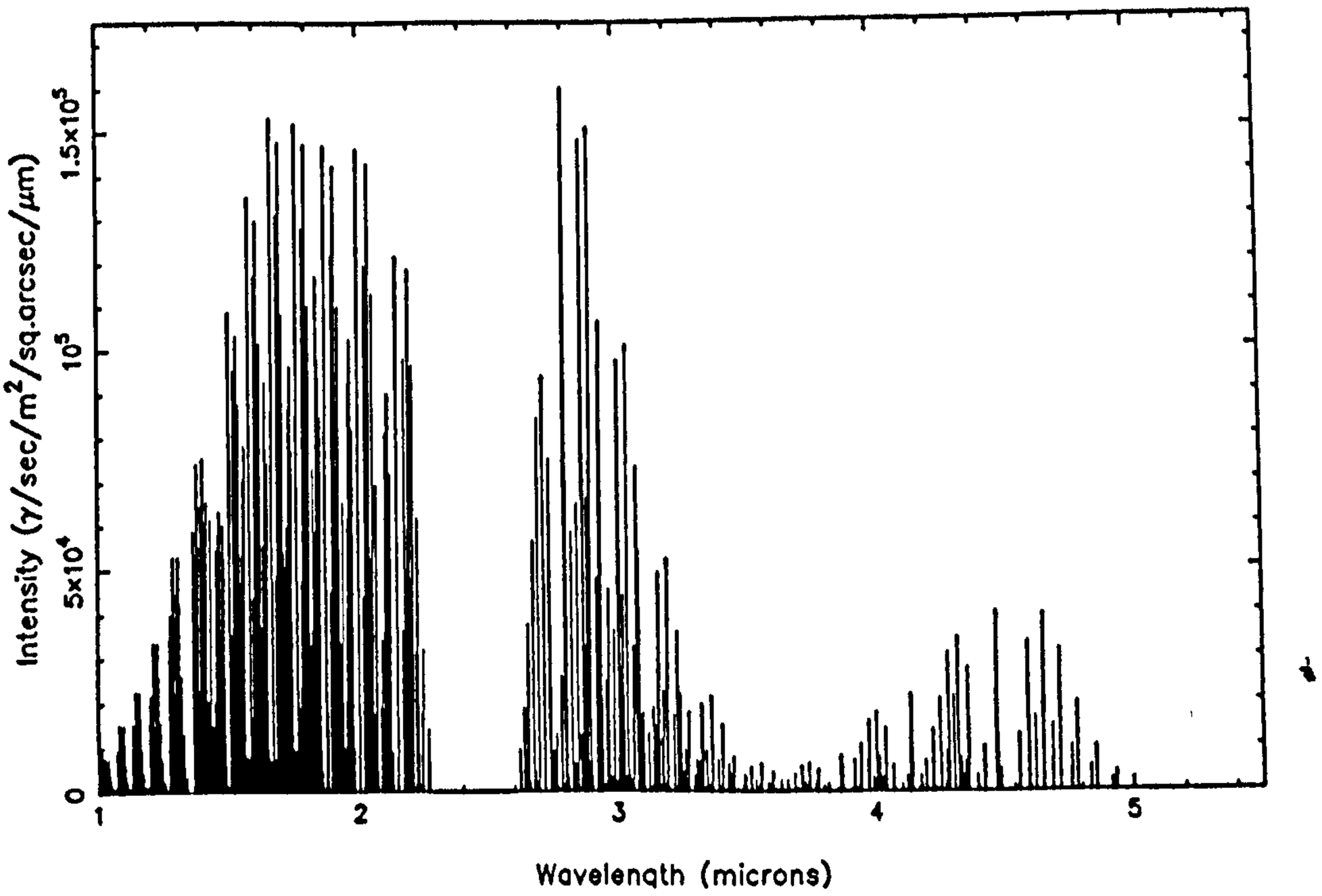


Figure 4.7: Intensity spectrum of OH* emission used by SIRCAM

metre per square arcsecond — the conversion is :

$$\mathrm{I} \quad (\text{in } \gamma/\mathrm{s}/\mathrm{m}^2/\widehat{\pi}^2) \quad = \quad 0.0187\ \mathrm{R} \tag{4.8}$$

Figure 4.7 shows the derived intensity spectrum of the OH* emission. This is in terms of the number of incoming photons above the lower atmosphere, and is referred to the zenith. Next we need to examine the various ways in which the actual photon fluxes may vary, as it is changes in the emission that limit may observational accuracy, rather than the actual level of the emission itself.

4.5.5.1 Zenith angle dependence

The OH* emission arises in a thin ($\sim 10\,\mathrm{km}$) layer at about 90 km altitude [88], and increases in intensity with zenith angle, as the telescope looks through a longer path length in the emitting layer. The form of the increase with zenith angle is given by the van Rhijn function, V [123, page 77] :

$$V = \left(\sqrt{1 - \left(\frac{r}{r+h}\right)^2 \sin^2 z} \right)^{-1} \tag{4.9}$$

where r is the radius of the Earth (6.4×10^6 m)

h is the height of the emitting layer above the Earth's surface

z is the zenith angle

For zenith angles of $\leq 45°$ and emitting layers with altitudes of ≤ 100 km, V is approximated by the airmass, sec(z), but at greater zenith angles and/or higher altitudes, the value of the van Rhijn function is slightly less than the airmass.

4.5.5.2 Altitude of site

As the emission arises in a layer about 90 km high, there is no decrease in this source of background at the higher altitude ground based sites. In fact, the total observed photon emission will probably be greater at these higher sites, as the tropospheric attenuation is decreased. This higher background might to indicate a lower final signal-to-noise at first sight, but recall that the error introduced by background subtraction increases only as the square root of the background. Thus if the atmospheric attenuation is halved at a higher site, then twice as many background *and* astronomical photons are detected—the resulting signal-to-noise *increases* by a factor of $\sqrt{2}$.

4.5.5.3 Geographical location

The mean intensities of some optical nightglow emission lines (e.g. OI 6300Å and 6364Å) show some geographical variations, due in this example to the location of the geomagnetic equator—indeed, Mauna Kea is in a particularly bright OI location [123, pages 66–67]. However, the intensity of the OH* emission is apparently uncorrelated with geographical location, at least between 45° N and 45° S [132].

4.5.5.4 Seasonal variations

There is some evidence for a slight seasonal variation in the overall intensity level of the OH* emission, rising by up to 50% in the autumn [137]. Such effects are in general small, and also Mauna Kea is thought to lie in the zone of minimum amplitude.

4.5.5.5 Diurnal variations

The diurnal variations in the OH* emission are complex, but as a rough approximation, the mean daytime and nighttime levels can be considered equal. Changes are thought to be due to the solar semi-diurnal tides in the atmosphere, downward transport of atomic constituents as the atmosphere cools, and changes in relative atomic ratios as the sun sets and rises, altering the photo-dissociation rates [137]. Modelling of the diurnal variations is yet to fit the observations fully [108], and there also appear to be seasonal and geographical components to the diurnal variations (not to be confused with seasonal and geographical variations *per se* [157]) — the only features worth noting here are that there appears to be a sharp dip in emission intensity immediately after sunset, rising back quickly, before falling slowly to a 'pre-midnight dip', recovering around 2am, and then falling slowly away again towards dawn. Such variations are small relative to those discussed next.

4.5.5.6 Other temporal variations

The most important temporal variability in the level of the OH* emission intensity appears to be due to the motion of gravitational waves in the upper atmosphere. Evidence for wave structure in the large scale OH* emission has been shown photographically [107], and more recently with intensified TV cameras [140], in the optical far red. These formations have similar structure to noctilucent clouds which also arise in the mesopause — the latter are only infrequently seen, whereas OH* wave structures will appear on most nights.

The wave systems are quite complex in structure, with parameters anywhere in the following ranges : total extent up to $10^6\,km^2$; length of individual waves up to 600 km; crest-to-crest wavelength between 5 and and a few hundred kilometres; peak-to-trough intensity amplitude ratio as high as 2; horizontal velocity from 10–100 ms^{-1}, with wave phase velocity and local wind velocity superimposed. Several such wave systems with different extents, wavelengths, and velocities can be superimposed simultaneously, leading to a highly spatially and temporally variable pattern. The source of the perturbations in the mesopause which leads to the propagation of these internal acoustic gravity waves is as yet unknown, but may be linked to solar and lunar tides in the atmosphere,

changing meteorological conditions, or auroral activity [136].

Waves of 50 km wavelength moving across the field of view at $15\,ms^{-1}$ can mean a factor of two change in the observed OH* emission intensity in half an hour — this is consistent with some reports of observed near-infrared background changes at Kitt Peak (Rieke, private communication). Others have reported very stable near-infrared backgrounds, which may also be possible, as OH* structure is not present every night. Also, there is probably a smaller scale structure variability within the individual waves, as yet unobserved. These variations will be relevant to astronomical observations with typical fields of view of one arcminute or less.

4.5.5.7 Spectral variability

Not too important for broad-band imaging observations, but relevant to high resolution near-infrared spectroscopy, is the variability in the spectral structure of the OH* emission. Changes in relative line and band intensities will limit the accuracy of removal of telluric contaminations from astronomical spectra. Such changes occur as the rotational temperature of the OH* varies. Also, the relative rotational temperature from branch to branch is not well known, and only the lowest level and brightest P branch lines are thought to have a Boltzmann distribution, the upper levels not being in thermal equilibrium with the local atmosphere [81].

All these factors make for complicated spectral variations. Within a given band, the rotational temperature may change by as much as 50 K in as little as a few hours, i.e. the timescale typical of gravitational wave motions [32], and changes are known to occur on all timescales, from minutes to years. A 50 K change in the rotational temperature of any given band can alter the absolute intensity of the higher level lines by as much as 100%, as the level population distribution is changed. Even the lower level lines can change by 50%, and the line-to-line intensity ratios can change by anywhere in the range ±100% [82].

4.5.5.8 Overall variability

As we have seen, the intensity of the OH* emission can be highly variable on timescales of minutes to years, with several factors of two variation possible, both temporally and spectrally. Both types of variation are likely to be important to broad-band imaging, and particularly so to near-infrared high resolution spectroscopy with integrating detectors, where the on-chip integration times may be very long [151].

4.5.6 O_2 IR atmospheric bands

The O_2 IR atmospheric bands are relatively limited in spectral extent and intensity compared to emissions from the OH* molecule, but are worth including here for their potential impact on the J broad-band filter background.

There are two main bands radiated by the O_2 molecule in the near infrared — the (0,0) and (0,1) bands at 1.27 μm and 1.58 μm respectively, arising as molecular oxygen drops from the $a^1\Delta_g$ level to the $X^3\Sigma_g^-$ level. There are three main mechanisms for creating excited O_2 in the $^1\Delta_g$ level — the first is by photo-dissociation of the ozone molecule and is only applicable during daylight or twilight, with the other two responsible for the nightglow. The second mechanism is due to the recombination of atomic oxygen in the presence of any molecular or atomic third body M [123, page 59] :

$$O + O + M \longrightarrow O_2(^1\Delta_g) + M \tag{4.10}$$

The observed covariance of O_2 and OH* emissions lead to the recognition of a third mechanism, namely the reaction between vibrationally excited OH* molecules and atomic oxygen [89] :

$$OH^* + O \longrightarrow H + O_2(^1\Delta_g) \tag{4.11}$$

The total intensity of the two bands is about 100 kR, with the 1.27 μm band being some 50 times brighter than the 1.58 μm band [147]. The observed ground level ratio will be much closer however, as the 1.27 μm band is heavily self-absorbed by O_2 molecules in the lower atmosphere [43]. A problem arises here due to this self-absorption. The SIRCAM model works at a maximum spectral resolution of 10^{-3} μm, and at this resolution the attenuation of the 1.27 μm band is poorly simulated. Using a single 10^{-3} μm wide bin at 1.27 μm containing the 100 kR flux predicted for the line, and passing it through

the smoothed absorption spectrum, greatly overestimates the amount of flux arriving above the telescope. Even working at higher spectral resolution would not answer all the uncertainties, as the rotational temperature of the O_2 molecules responsible for the emission in the mesopause will be different to that for the molecules responsible for the absorption lower in the troposphere, and the relative contributions of the various components of the R and P branch lines would be difficult to calculate.

The effort required to adapt SIRCAM to the very high spectral resolution required to solve this problem is not acceptable, considering that IRCAM is in general an imaging system, rather than a spectrometer, and therefore very high spectral resolution simulations are largely unnecessary. However, for two reasons, it is important to input an approximate solution into SIRCAM. Firstly, the potential impact on the J broad band filter is large, given the high intrinsic brightness of the 1.27 μm band above the atmosphere, and the low OH* brightness and non-existent thermal contributions seen through that filter. Secondly, in the imaging spectroscopy mode, IRCAM uses a Fabry-Pérot étalon to achieve spectral resolutions $\sim$ 1000, corresponding to $\sim 10^{-3}$ μm at J, at which point the spectral structure of the O_2 bands may become of concern.

Examining a ground-level airglow spectrum showing the O_2 1.27 μm band along with several of the OH* bands [147], the 1.27 μm band appears to be as bright as the Q branch of the OH* (8,6) band. The simulated spectrum shown in section 4.5.5 above gives a brightness of this OH* Q branch of about 20 kR, indicating that the O_2 1.27 μm band is attenuated by about a factor of five in the lower atmosphere. This is consistent with other reports of the ground-level brightness of the 1.27 μm band [43]. Thus, in the SIRCAM model, the 1.27 μm band is artificially attenuated by a factor of 4.5 above the atmosphere, to a brightness of 22 kR, with the subsequent atmospheric transmission of about 0.9 giving an overall attenuation of five. The 1.58 μm band is left unaltered at a level of 2 kR, as it is not self-absorbed. Note that this issue of self-absorption does not apply to the OH* emission discussed in section 4.5.5, as there is little or no OH in the lower atmosphere.

The O_2 IR atmospheric bands arise in the mesopause alongside the OH* bands, with which they show some correlation as mentioned above. Thus, the previous discussion on intensity variability of the OH* bands is assumed to roughly apply to the O_2 bands also. In fact, the intensity of the O_2 bands has been shown to vary by as much as a

factor of six in a few nights [115].

4.5.7 The nightglow continuum

For completeness, we include the so-called 'nightglow continuum', even though its contribution to the total intensity is small for most applications. Both the nature and intensity of the near-infrared continuum are poorly known — the reaction of $NO+O \longrightarrow NO_2+h\nu$ responsible for the observed continuum shortward of 7000Å [28, pages 543–545], extends at low level into the near-infrared, and radiative association of atomic oxygen to form vibrationally excited O_2 has been suggested to explain a possible rise of the continuum into the infrared [163]. Subsequent observations have demonstrated no such rise, and have the left the origin and intensity of the continuum in doubt [115,130].

The main reason for the confusion is the low level of the emission and difficulties in removing contamination by the OH* lines. We shall adopt a fairly low and constant value of 15,000 R/μm throughout the near-infrared [130], noting that this translates to only 3000 γ/s/$\hat{\pi}^2$/μm on the UKIRT, meaning it will be notable only at J, or at very high spectral resolutions elsewhere.

Variability is not of great importance given the low intensity, but factor two changes in intensity can be seen in less than half an hour, again on the same timescale as the gravitational wave motions [130]. As noted above, some of the possible mechanisms might indicate covariance with the intensity of the OH* and O_2 emissions.

4.5.8 Total non-thermal emission

The total non-thermal emission spectrum is calculated by adding the OH*, O_2, and infrared continuum spectra. The result is completely dominated by the OH* emission, with the only noticeable addition being the single bright O_2 line at 1.27 μm. It is worth noting that there are 'fudge factors' built into SIRCAM which allow each component of the nightglow spectrum to be scaled by a different constant, in order to simulate the possible temporal variations. No simulation of the spectral structure changes is made at this time.

4.5.8.1 Comparison with actual sky background data

Now that we have derived a model of the near infrared emission spectrum, we should like to compare it with actual measurements of the sky background. However, such measurements of the absolute sky background flux seen at astronomical observing sites between 1 and 2.5 μm have been rare. The simple reason for this was the lack of sensitive DC coupled instruments. The chopping technique used with AC coupled detectors removes the absolute sky level, leaving only the source flux and sky noise. Some single element detectors have been used in a DC coupled fashion to measure the background, but only as an integrated brightness through a broad band filter [50]. In fact, the poor understanding of the spatial and spectral structure and variability of the near infrared sky background was one of the main motivations of this work, as these parameters will become increasingly important as infrared integrating arrays are applied to imaging, and particularly spectroscopy.

A low resolution spectrum of the near infrared sky brightness above Mauna Kea has been obtained using a DC coupled single element detector and a CVF on the IRTF (Tokunaga, private communication). However, this spectrum suffers from low spectral resolution ($\sim$ 3%), and contamination shortward of 1.6 μm by lunar scattering, and longward of 2.3 μm due to poor removal of the steeply rising thermal background from the warm optics, making accurate comparison between it and the model data difficult. More useful is a recent spectrum obtained by Alan Moorwood at ESO in Chile, using the integrating array spectrometer, IRSPEC [106]. The spectrum is reprinted in figure 4.8 with the permission of Alan Moorwood and shows a portion of the H window at a resolution of approximately 1000, and was obtained with a $6'' \times 6''$ aperture on the ESO 3.6 m telescope pointed at the zenith. Instrumental offsets were cancelled by observing with the spectrometer slit open and closed, and flux calibration was obtained by observing a photometric standard star, and assuming a smooth energy distribution for the star in the near infrared (Moorwood, private communication).

For comparison, we have smoothed the equivalent wavelength range of the model non-thermal emission spectrum derived above, to the resolution of Moorwood's spectrum, converting the flux into matching units of Rayleighs per micron. This is shown in figure 4.9. We can immediately see that the correspondence between the observed and model spectra is very good, both in the observed intensities and spectral structure.

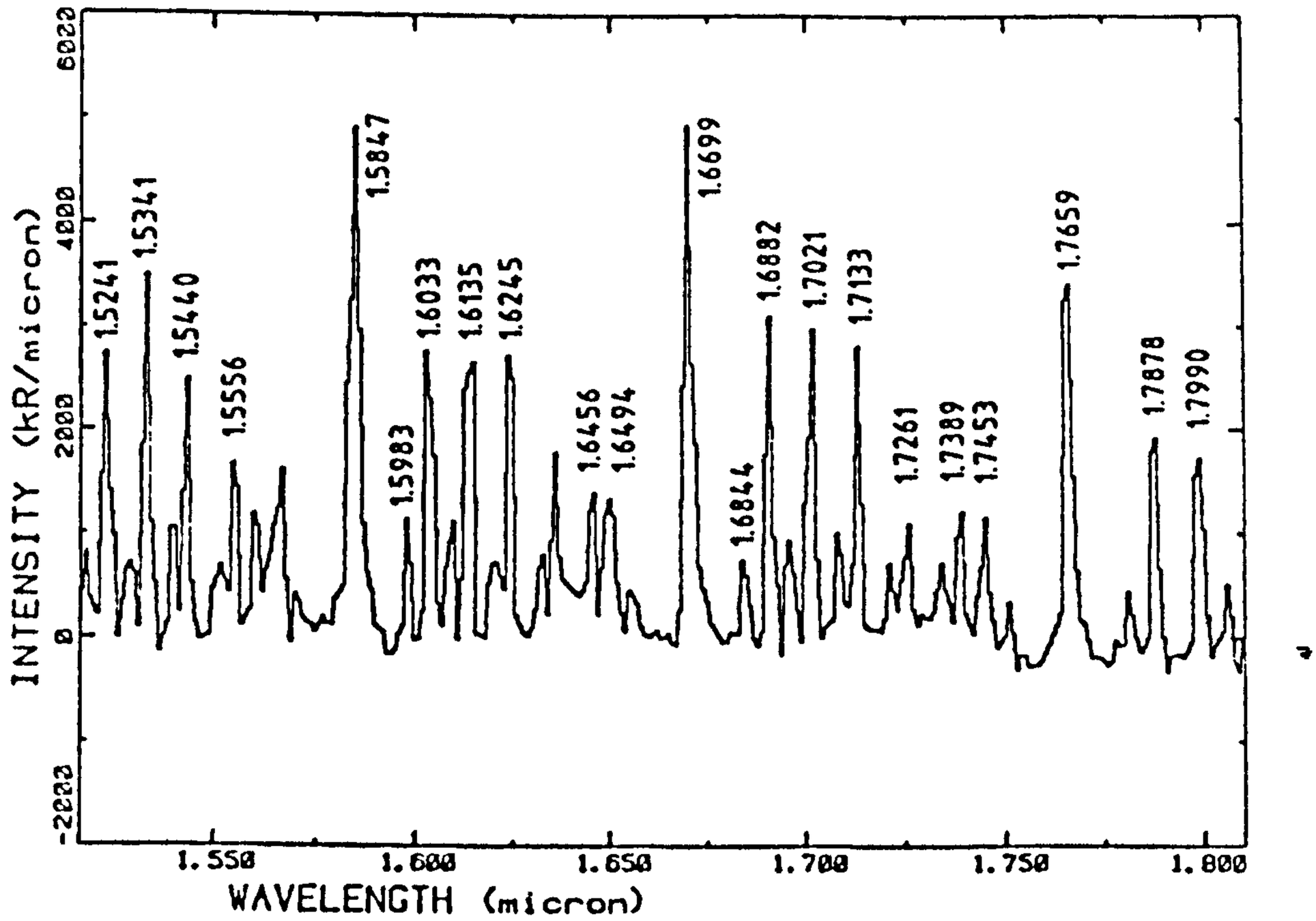


Figure 4.8: IRSPEC spectrum of the near infrared sky background

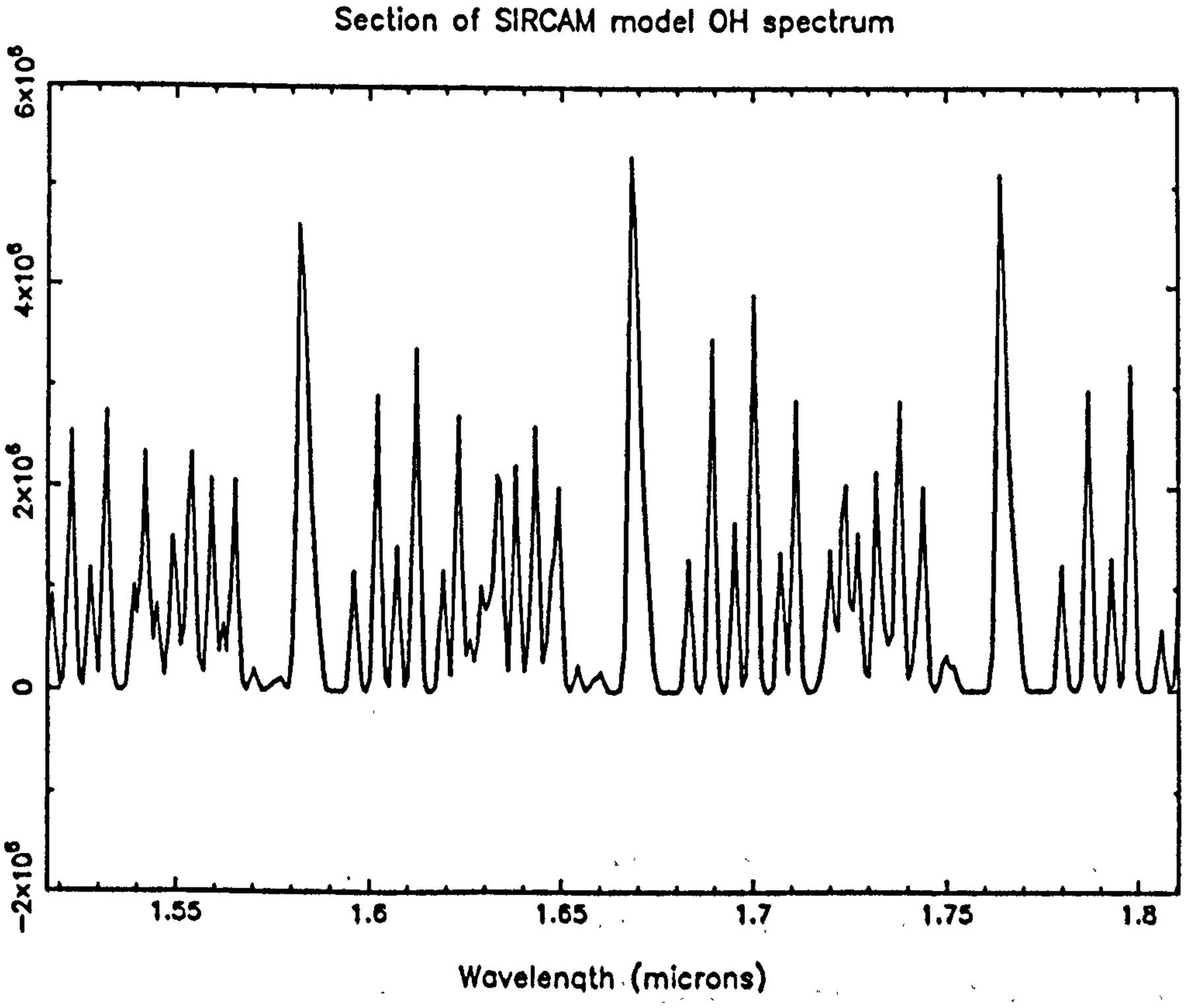


Figure 4.9: Comparison SIRCAM model spectrum

In detail, some line ratios are somewhat different, and several lines appear to be missing from the model spectrum. However, these differences may easily be accounted for by considering the complexity of the emission processes discussed above, if the rotational temperature used for the model was different to that of the mesopause at the time the actual data were obtained, for example.

4.5.8.2 Conclusions

Although this comparison examines only a limited portion of the near infrared sky background, it tends to suggest that it is well modelled by this work. There are discrepancies between the model used here and work done by others [135], but this is mainly confined to the higher Δv transitions, which give rise to emission shortward of $1\,\mu$m (Llewellyn, private communication). Thus, bearing in mind that the previous sections on temporal, spectral, and spatial variability show that factors of two change in absolute intensity are easily possible, the model non-thermal emission spectrum derived here may be used to derive a reasonably accurate indication of the near infrared sky background that may be expected between 1 and $2.5\,\mu$m at low and high spectral resolutions.

4.5.9 Lunar scattering

The moon is effectively a low albedo mirror for the sun, and illuminates the night sky with a colour temperature of 5900 K, peaking at about $0.5\,\mu$m. Therefore, in the infrared, the lunar spectrum follows the Rayleigh-Jeans approximation, falling off as λ^{-4}. This flux may be scattered into the beam, and be detected. At the altitude of Mauna Kea, the scattering is due to molecules and high stratospheric aerosols [161, page 3-61], both small compared to the wavelength of near-infrared radiation—it is valid to assume Rayleigh scattering, again with a λ^{-4} dependence. Thus the cumulative wavelength dependence of the sky brightness will be as λ^{-8}, falling off sharply into the near-infrared, and the only filter significantly affected by lunar scattering is the J filter at $1.2\,\mu$m. Obviously, the total amount of flux detected will depend on the lunar phase, the altitude of the moon, and the angular distance of the moon from the point in the sky being imaged. As these parameters are very variable, lunar scattering has been excluded from the SIRCAM model.

4.6 Telescope and warm optics

In this section we shall discuss the rôle played by the telescope and other warm optics. These components sit at ambient temperature and are responsible for focussing the incoming radiation (astronomical photons and the sky background) into the cryogenic instrument. This radiation is attenuated by the imperfect reflections and transmissions in this optical train, and also, a further thermal background is added as the warm optics radiate black-body photons into the beam. Three important parameters are associated with the warm optics :

- the reflectivity or transmission of each component, which determines the fraction of the photons arriving above the telescope which reach the cold instrument
- the emissivity of each component, which determines the black-body radiation added to the beam
- the temperature of the warm optics, the other parameter that determines the amount of black-body radiation

The reflectivities, transmissions, and emissivities are treated spectrally, as these parameters may vary according to wavelength, depending on the materials used. Also, the question of 'dirt' is addressed; that is, the lowering of reflectivity or transmission and corresponding rise in emissivity due to oxidation, tarnishing, and dust contamination.

The SIRCAM model allows many warm optical configurations to be simulated, but we shall concentrate here on the optical train relevant to this thesis, namely IRCAM on the UKIRT on Mauna Kea. The UKIRT operates as a Cassegrain reflector, with primary and secondary mirrors. A tertiary dichroic mirror folds the beam through 90°, after which two gold mirrors fold and collimate the beam prior to entering the IRCAM cryostat through an entrance window.

4.6.1 Primary

The UKIRT primary is a thin parabolic Cervit mirror coated with an aluminium film. The nominal diameter is 3.8 metres, with an effective diameter of 3.699 m, the reduction

due to the secondary 'underfilling' the primary. That is, whilst conventional optical telescopes have secondary mirrors somewhat bigger than needed to intercept the converging beam from the primary in order not to waste any astronomical photons, an infrared telescope has a secondary mirror smaller than the size of the primary beam at the intercept point. This is to prevent the secondary viewing the support structure and floor beyond the edge of the primary — these are highly emissive room temperature sources which would raise considerably the thermal background seen by the instrument. The central hole in the primary has a diameter of 1.023 m, and thus the effective collecting area of the primary, A_{tel}, is 9.924 m^2. This area along with the solid angle of sky imaged by each pixel, Ω_{pix}, forms the basic $A\Omega$ product for the system, a conserved quantity. Thus, although it is in fact the underfilled secondary that is the effective system stop, we shall use $A_{tel}\Omega_{pix}$ as the $A\Omega$ product for all optical components in the system.

The spectral reflectivity $R_p(\lambda)$ of clean aluminium is shown in figure 4.10 [161, page 7-81]. The UKIRT primary is re-aluminised annually and washed regularly, but nevertheless the primary is probably the dirtiest component in the system due to oxidation of the aluminium and its exposed up-facing position making it prone to dust collection, for which Mauna Kea is well-known. As a first approximation, this 'dirt' is assumed to be grey, i.e. that the attenuation due to the oxidation and dust has no spectral dependence. Multiplying the spectrum shown in figure 4.10 by a constant 'dirt factor' between 0 and 1 simulates this effect. A typical value used for the primary dirt factor (PDF) is 0.95. This approach is used for the following components also. The dirt factors used throughout the modelling have been taken from a previous study, in which SIRCAM was used to simulate and analyse the emissivity of the UKIRT as measured with a single channel photometer [94]. It should however be noted that the 'dirt' is probably not grey, but that it exhibits a wavelength dependence (Lee, private communication).

The imperfect reflectivity is due to absorption by the aluminium and the dust on the aluminium, and thus the emissivity of the primary may be found from Kirchoff's law, by taking one minus the (dirty) spectral reflectivity.

4.6.2 Secondary

The hyperbolic secondary mirror is also aluminium coated glass, with a spectral reflectivity $R_s(\lambda)$ as for the primary. Due to its downwards facing position however, the secondary stays very clean, and the dirt factor for the secondary (SDF) will be close to unity. Again, the emissivity is found via Kirchoff's law.

Two potential problems arise at the secondary. Firstly, the central hole in the primary mirror is seen as a highly emissive source of thermal background by the secondary. To eliminate this extra background, a small off-axis flat mirror replaces the central section of the hyperbolic secondary, such that a clear area of the primary and thence the sky is seen, rather than the central hole. Secondly, the support structure which holds the secondary in place, known as the spider, must have a minimal cross-sectional area in order to minimise its contribution to the thermal background, as it lies in the beam as seen by the primary. This is so for the UKIRT, and it can be shown that a simple cross-shaped spider with a cross-section of 1 cm and emissivity of 1, will increase the total thermal flux emitted by the warm optics by about only 3%, and thus has been neglected in the SIRCAM model. It is true in general that infrared telescope designs are mechanically very open, such that minimal additional thermal background radiation is added.

4.6.3 Dichroic

The UKIRT design employs a 45° incidence mirror in the converging beam just above the Cassegrain focus, in order to fold the beam through 90° and out to a side-looking instrument mounted on the ISU2 instrument support unit. By using a calculated thickness of gold film on a glass substrate, a beamsplitting or dichroic mirror is formed, with the optical radiation passing straight through a guide camera mounted at the Cassegrain focus, and the infrared radiation folded out to the bent Cassegrain focus as desired. The spectral reflectivity of the UKIRT dichroic mirror $R_d(\lambda)$ is shown in figure 4.11 (Lee, private communication).

The dichroic is washed or replaced quite frequently, but nevertheless picks up dust due to its upfacing position. A typical dirt factor (DDF) of 0.98 is used.

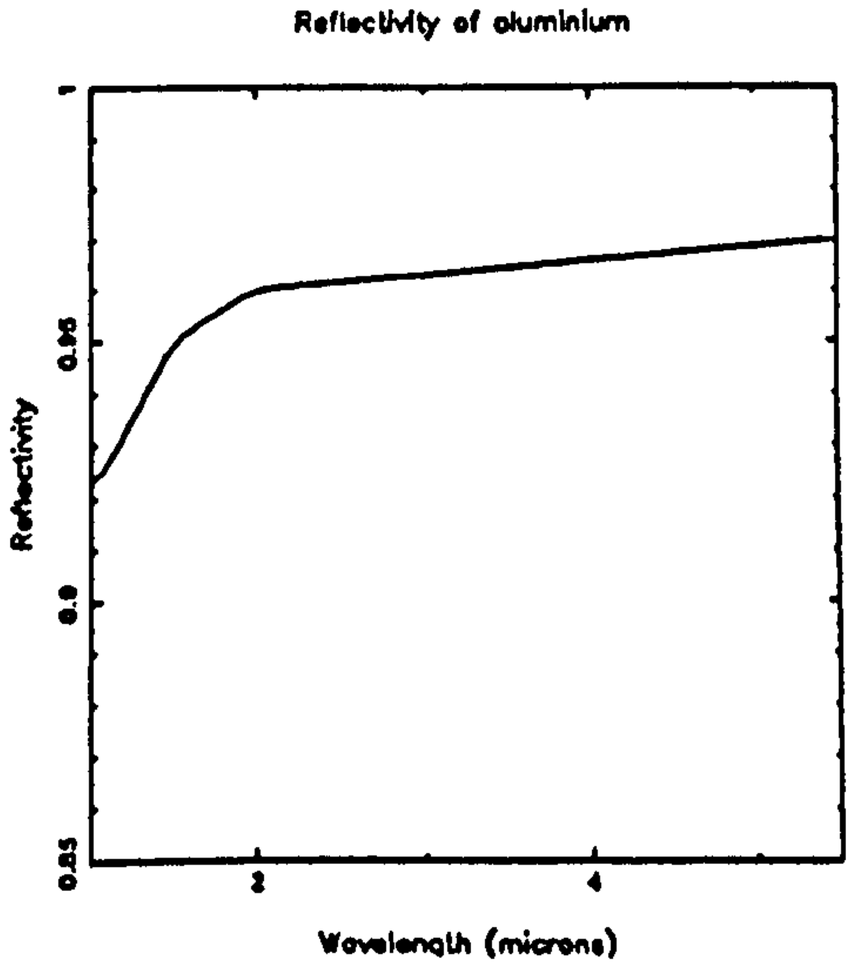


Figure 4.10: Reflectivity of aluminium

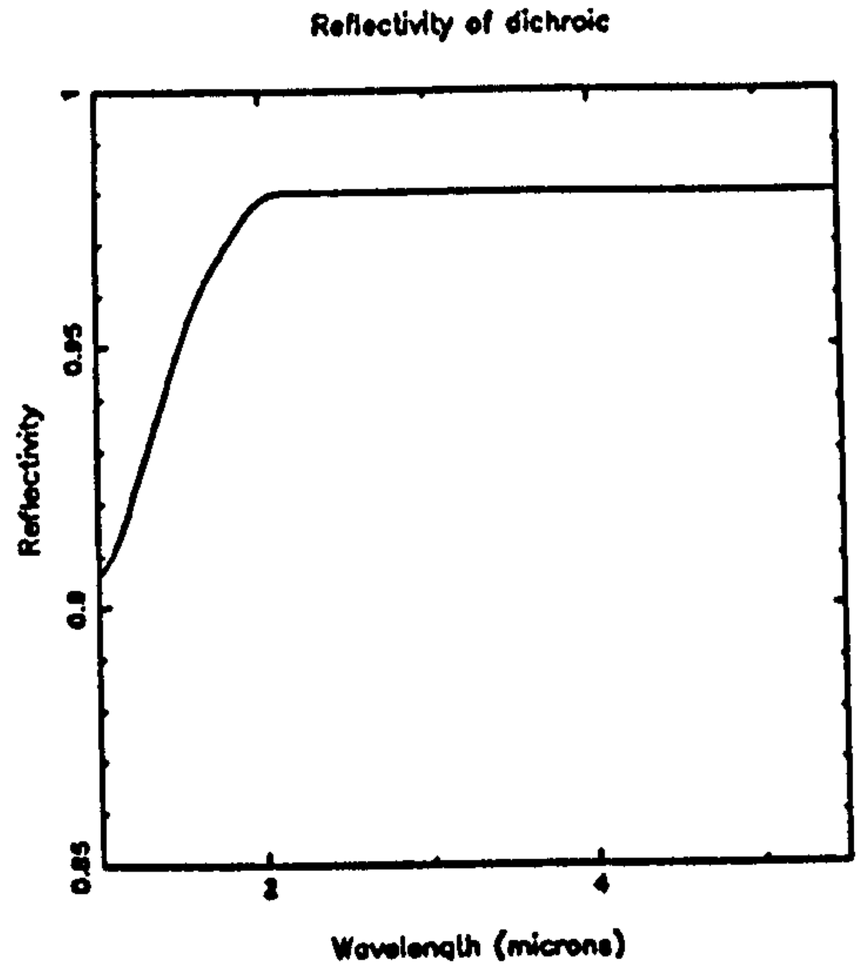


Figure 4.11: Reflectivity of dichroic

Kirchoff's law does not strictly apply to the dichroic, as the imperfect reflectivity is due to transmission losses as well as absorption losses at the shorter wavelengths. However, by 2.5 μm, where its thermal emission becomes appreciable, the dichroic is essentially a solid gold mirror, at which Kirchoff's law does hold. Thus, Kirchoff's law *is* used in order to find the emissivity of the dichroic.

4.6.4 Other warm optics

There are three additional warm optical components to be considered in the IRCAM design. Two gold mirrors are used to fold and collimate the beam prior to entering the cryostat through a calcium fluoride window. The reflectivity of gold $R_g(\lambda)$ is flat across the 1–5.5 μm region and is shown in figure 4.12 [161, page 7-81]. A typical dirt factor (GDF) of 0.99 is used for these side-looking, easily accessible and maintained mirrors. The emissivity is found via Kirchoff's law.

The entrance window of the cryostat is made of uncoated calcium fluoride (CaF_2) which has a *single surface* transmission of 0.97 [5] — the effective transmission $T_w(\lambda)$ for the 3 mm thick window is shown in figure 4.13. The window dirt factor (WDF) is set to unity, as the side-looking window is well protected from dust by the camera front baffle tube.

The emissivity of the window is essentially zero. Its transmission losses are due to reflection at the air-CaF_2 boundaries, not absorption in the CaF_2, and thus Kirchoff's law does not apply.

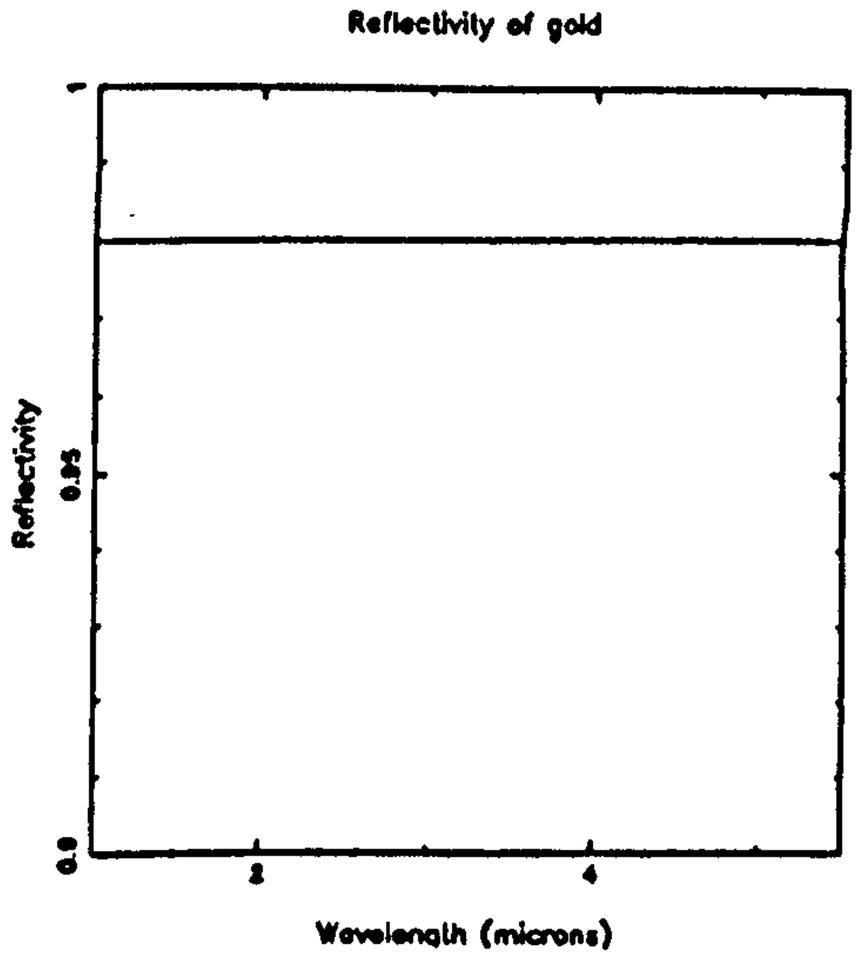


Figure 4.12: Reflectivity of gold

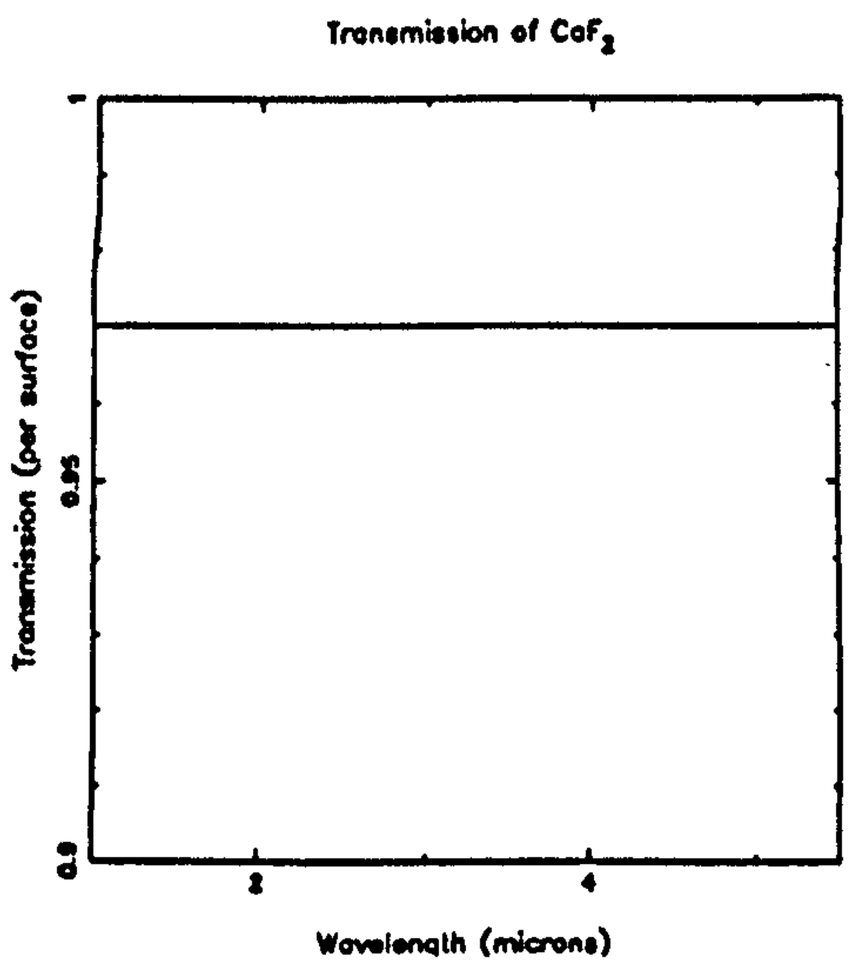


Figure 4.13: Transmission of CaF_2

4.6.5 Temperature of the warm optics

For simplicity, all the 'warm' optical components described are assumed to be at the same temperature, namely the ambient temperature of the dome. The ambient air temperature on Mauna Kea is low at around 0°C, with diurnal variations of 1–2°C, and only slight seasonal variations [109]. As long as the optics are shielded from the direct heating of the Sun, the thermal emission will be both low and stable, advantageous in reducing background shot and $1/f$ noise components. This is in contrast to lower altitude continental sites such as Kitt Peak, where diurnal variations can be 10–20°C, and the seasonal range can be as much as 40°C [4, page 5-11].

For the majority of SIRCAM simulations, Mauna Kea ambient temperatures of 0–2°C (273–275 K) are assumed for the telescope and warm optics.

4.6.6 Overall transmission and emissivity

The overall spectral transmission $T_{warm}(\lambda)$ for the warm optics train is found by multiplying together the individual reflectivity or transmission spectra together, each one having been multiplied by the relevant dirt factor.

$$T_{warm}(\lambda) = R_p(\lambda).\mathrm{PDF}.R_s(\lambda).\mathrm{SDF}.R_d(\lambda).\mathrm{DDF}.(R_g(\lambda).\mathrm{GDF})^2.T_w(\lambda).\mathrm{WDF} \quad (4.12)$$

The overall spectral emissivity $\varepsilon_{warm}(\lambda)$ is found by multiplying together the spectra and dirt factors for those components assumed to obey Kirchoff's law (i.e. all those

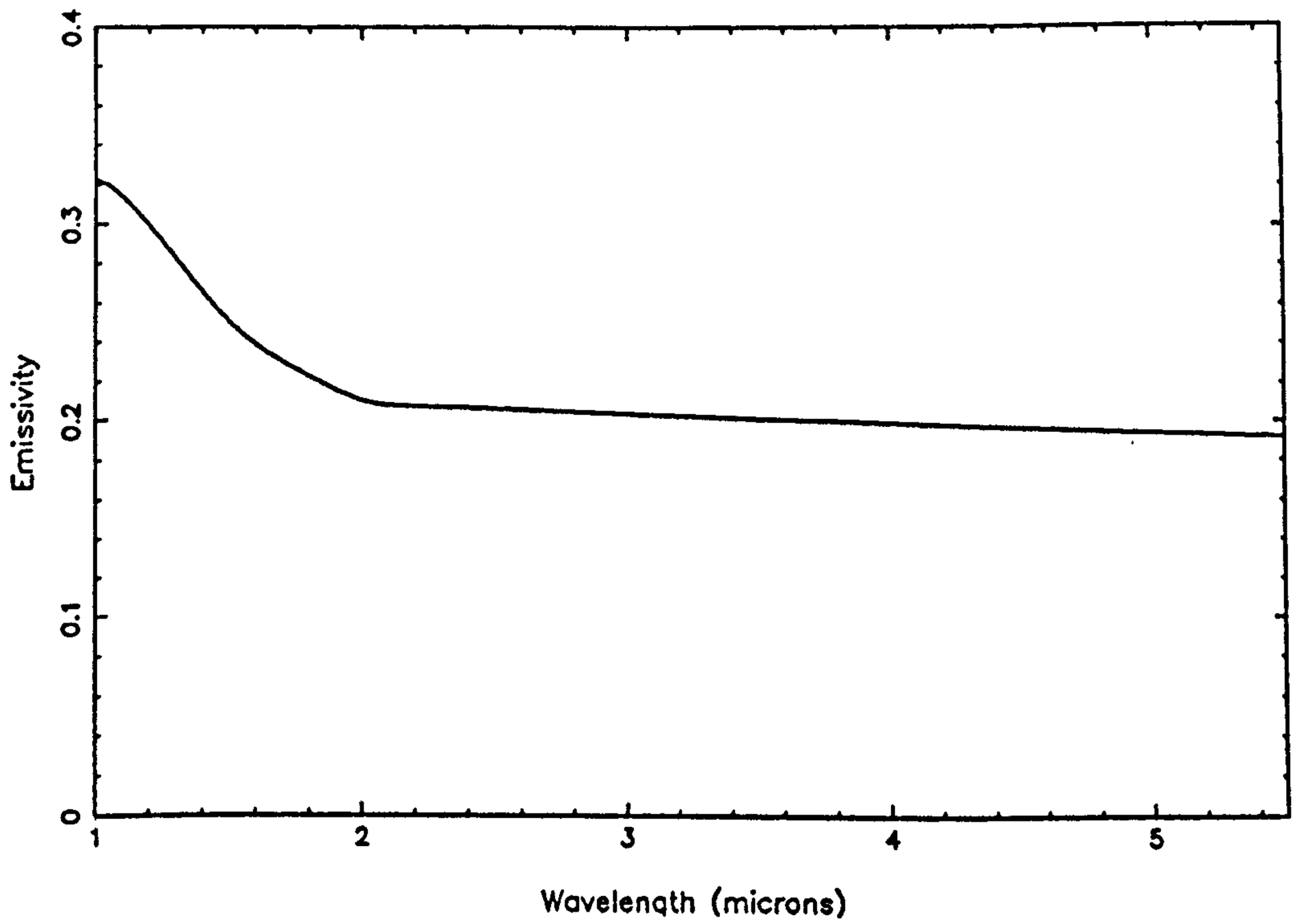


Figure 4.14: Emissivity profile for the UKIRT and IRCAM warm optical train

except the CaF_2 window), and then subtracting the resulting curve from one.

$$\epsilon_{warm}(\lambda) = 1 - (R_p(\lambda).\mathrm{PDF}.R_s(\lambda).\mathrm{SDF}.R_d(\lambda).\mathrm{DDF}.(R_g(\lambda).\mathrm{GDF})^2) \qquad (4.13)$$

From this equation, we can derive the spectral dependence of the warm emissivity. Figure 4.14 shows the spectral dependence of the emissivity of the warm optics $\epsilon_{warm}(\lambda)$ for the observational configuration discussed. This curve is important, as the shot noise on the thermal background emitted by the telescope and warm optics will mainly determine the performance of the camera at wavelengths longward of 2.5 μm. Note the turn up in emissivity at shorter wavelengths, due to the reduced reflectivity of the aluminium primary and secondary mirrors, and of the dichroic beam splitter. This is unimportant, as very few thermal photons will be emitted by the warm optics shortward of 2.5 μm.

4.6.7 Total warm background

We can now predict the spectrum of the total background flux, comprising of the non-thermal emission from the mesopause, thermal emission from the lower atmosphere, and the thermal emission from the warm optical components.

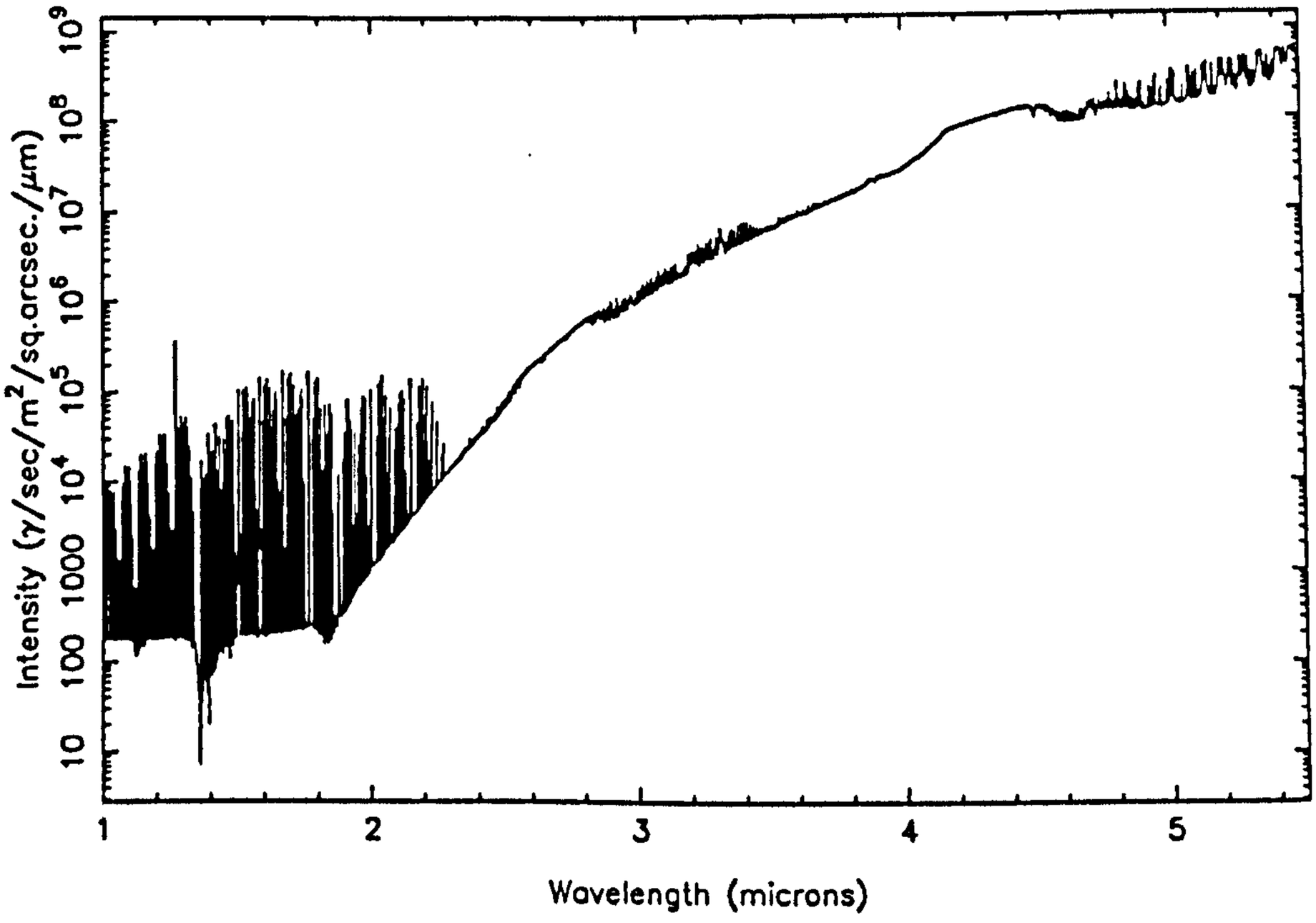


Figure 4.15: Background flux just inside cryostat window

We have calculated the spectrum seen just inside the CaF_2 window with the following parameters. We have assumed nominal levels of non-thermal emission, and a lower atmospheric temperature of 253 K. An airmass of 1.5, and a water vapour of 1.2 mm have been used. We have used the standard IRCAM warm optical train of primary, secondary, dichroic, gold flat, gold collimator, and CaF_2 front window. All these components have been given the dirt factors mentioned above, and the warm optical train is assumed to be at a temperature of 275 K. For the warm optical train, the emissivity was calculated as one minus the transmission, as discussed above.

The resulting spectrum is shown in figure 4.15. We can immediately see the two portions of the near infrared, the non-thermal and the thermal. The non-thermal, shortward of 2.5 μm, is characterised by a bright, spectrally highly structured background, whilst the background in the thermal section beyond 2.5 μm is much brighter still. It is characterised by the smooth blackbody function of the warm optics, superimposed with slight structure from atmospheric thermal emission. It is a spectrum such as this that now enters the cold instrument.

4.7 Camera

The optical components that definitely require cooling in an infrared instrument are the filters and all following optics. The reason for this is simple — a room temperature black-body curve at 273 K peaks at about 13 μm and thus near-infrared instruments (1–5.5 μm) are sensitive to the Wien part of this black-body. The steep exponential rise of the Wien section means that the largest numbers of photons will be emitted near the detector long wavelength cut-off. Thus most of these room temperature black-body photons will be filtered out by (say) a 2.2 μm filter, but then the filter must be cooled such that the long wavelength photons are not just re-emitted by the side of the filter facing the detector. Similarly, all optical components which follow the filter must be cold enough not to radiate more than a very small number of photons at the long wavelength cut-off of the detector material. It is easy to show that, in order to reduce the total number of black-body photons emitted by the internal camera optics to levels below typical InSb dark currents, the optics must be cooled to $\sim$ 80 K [120]. In this section, we assume that the cold optics (filters and lenses) merely attenuate the input beam, without addition of any extra thermal flux. Also, the whole internal cryogenic chamber is assumed to be cold enough not to add any thermal background either.

4.7.1 Filters

The cooled filters provide the wavelength selection. Both broad and narrow band interference filters are used in IRCAM, to provide spectral resolutions from 1–30%. In most cases the relevant broad filter is used to block a narrow band filter, as the latter may have leaks well away from the wavelength they were designed to pass.

The spectral transmission profiles for the standard OCLI near-infrared filter set (J,H,K,L′,M) are shown in figures 4.16 to 4.20 (Mountain, private communication). These profiles were measured at 77 K — the bandpass of interference filters moves to shorter wavelengths as the filter is cooled. This is due to the filter contracting at lower temperatures, reducing the path length between the interference layers, and thus decreasing the wavelength for which the filter acts constructively. A filter will typically move by about 2% shortward in wavelength when cooled from room temperature to liquid nitrogen temperature [3].

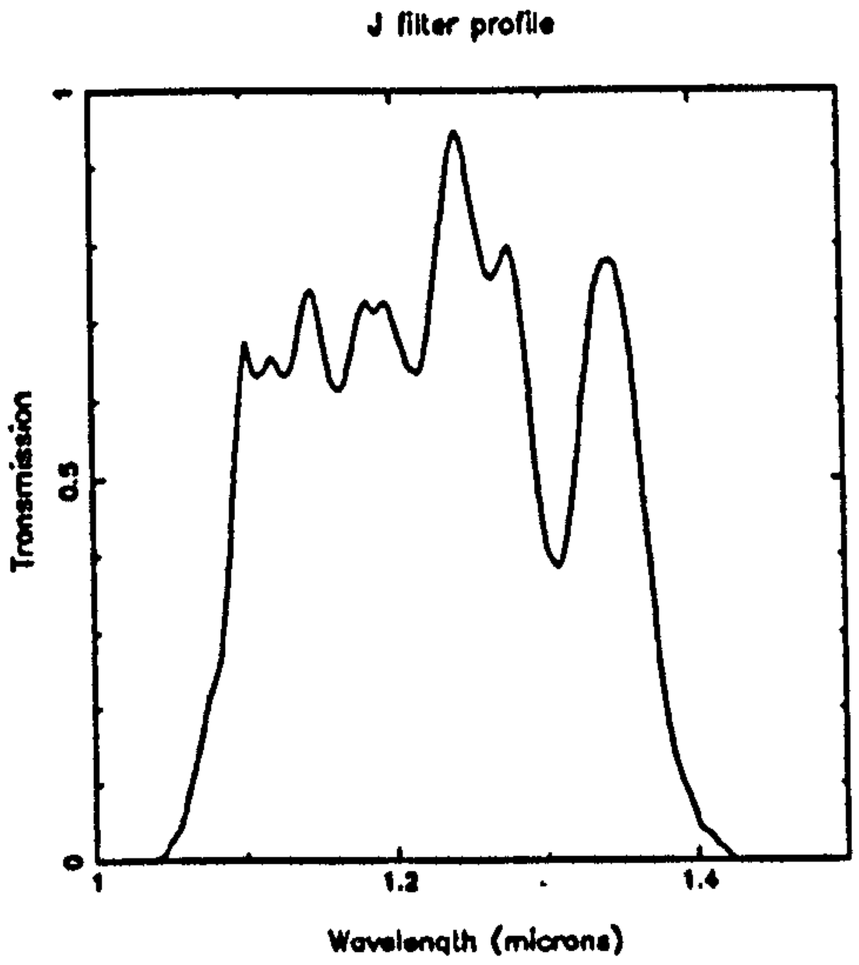


Figure 4.16: J filter profile

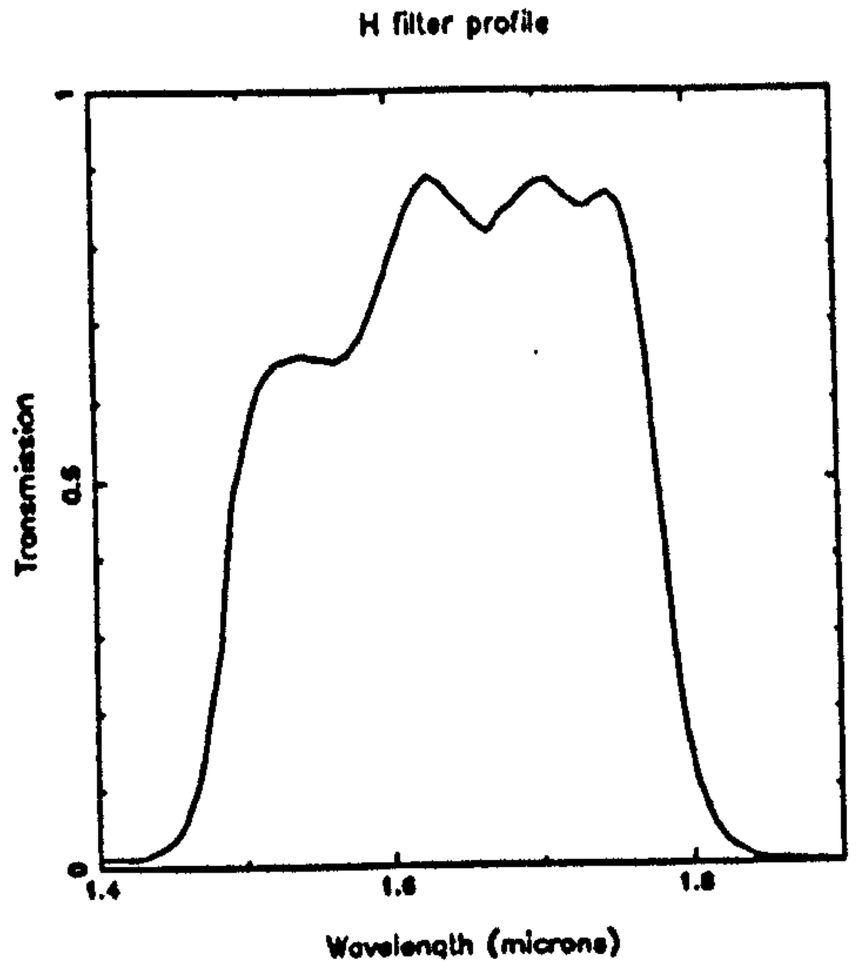


Figure 4.17: H filter profile

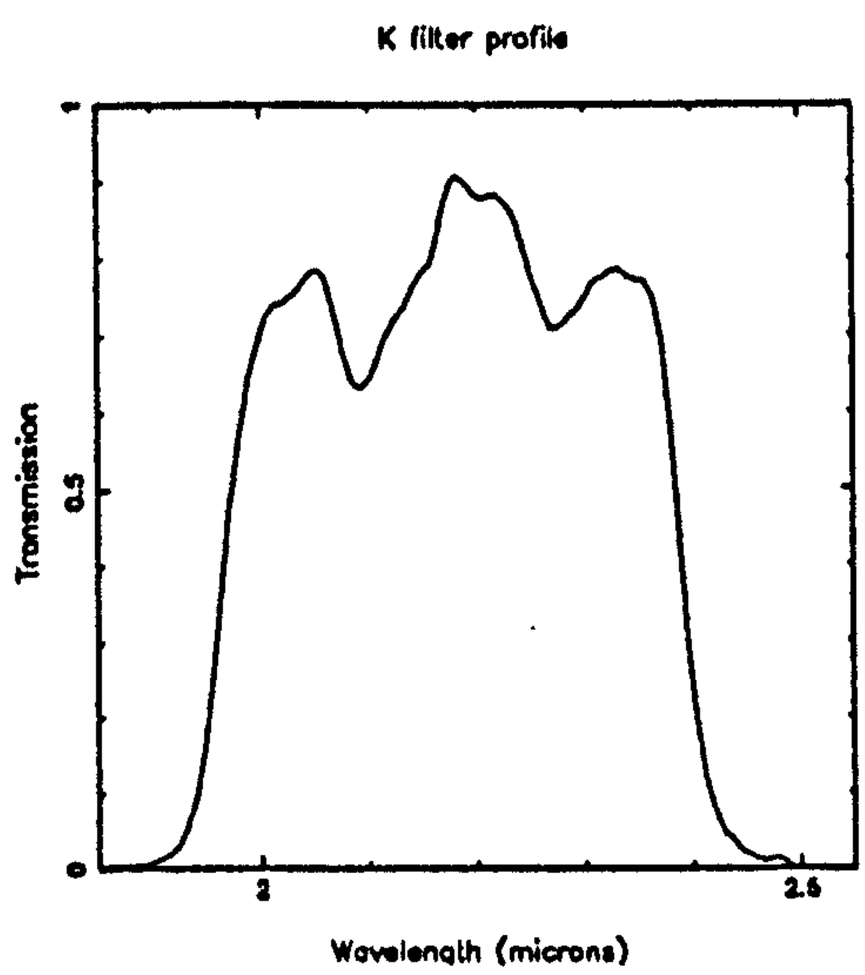


Figure 4.18: K filter profile

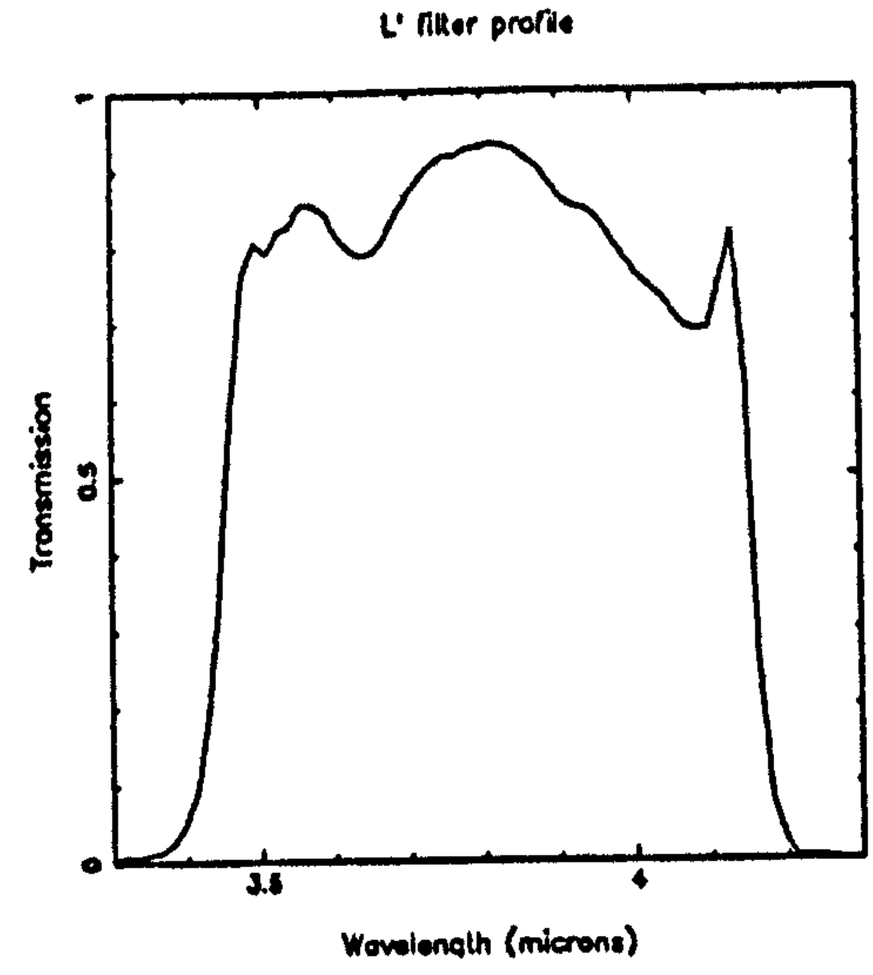


Figure 4.19: L′filter profile

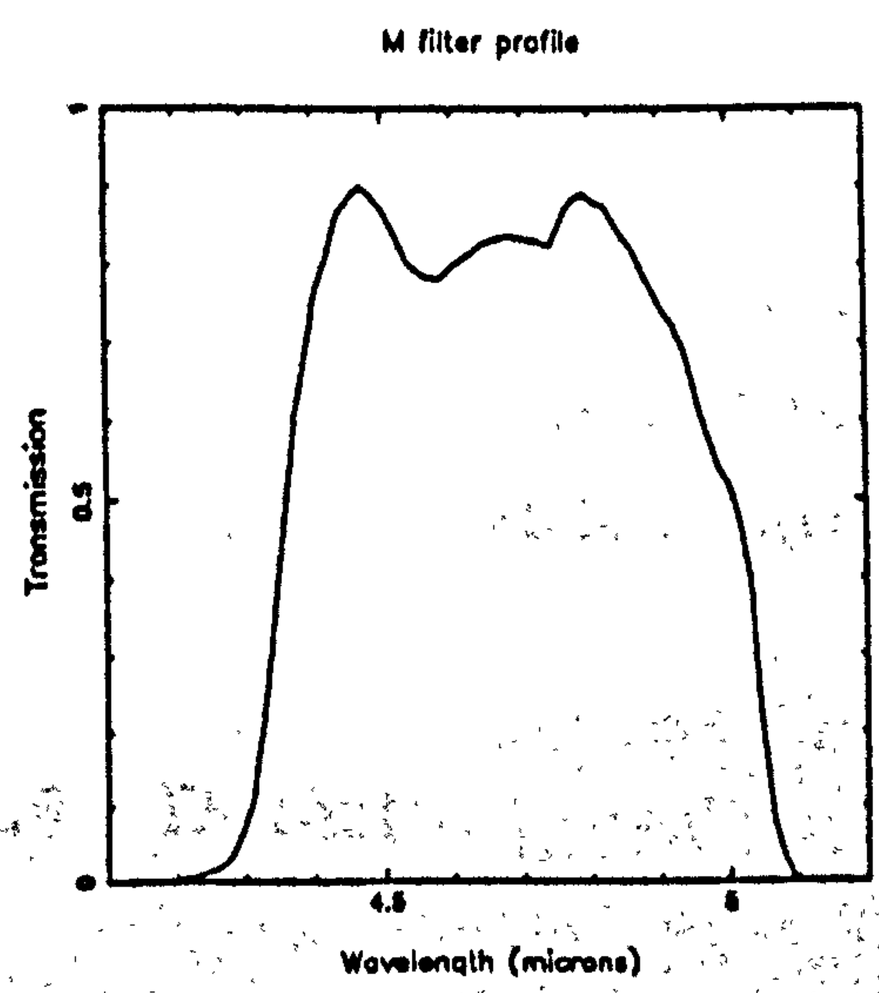


Figure 4.20: M filter profile

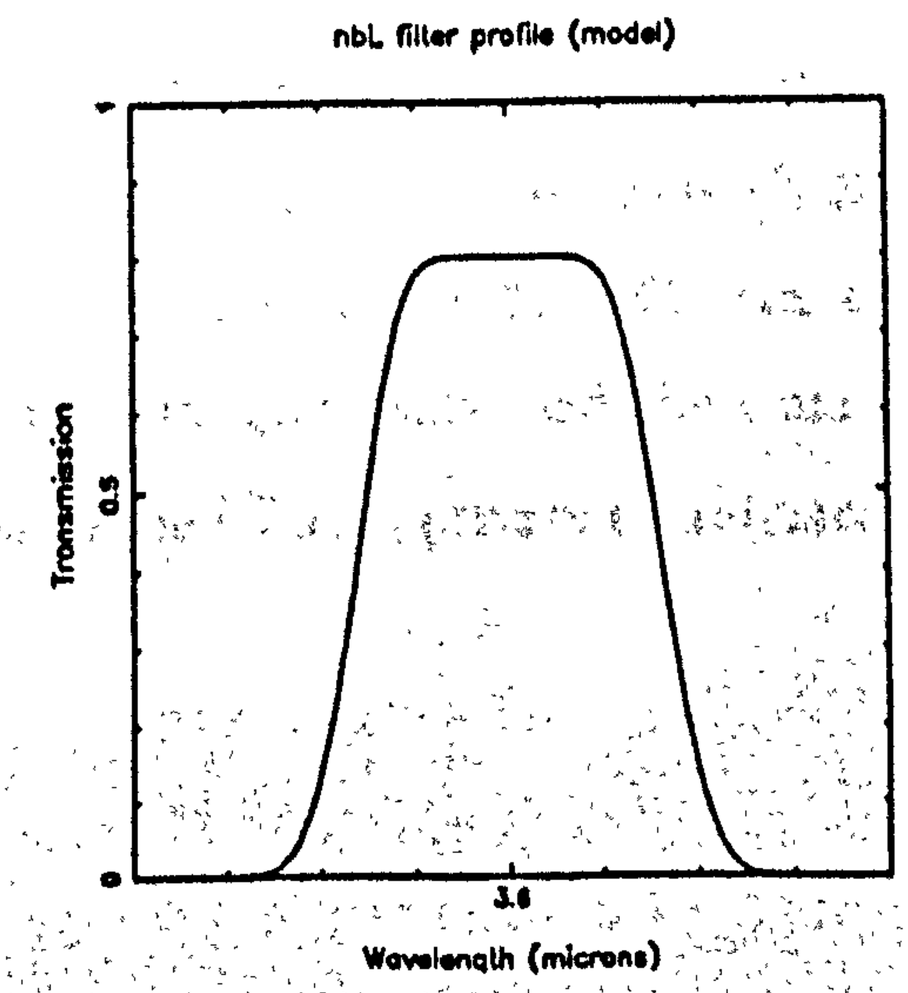


Figure 4.21: nbL filter profile

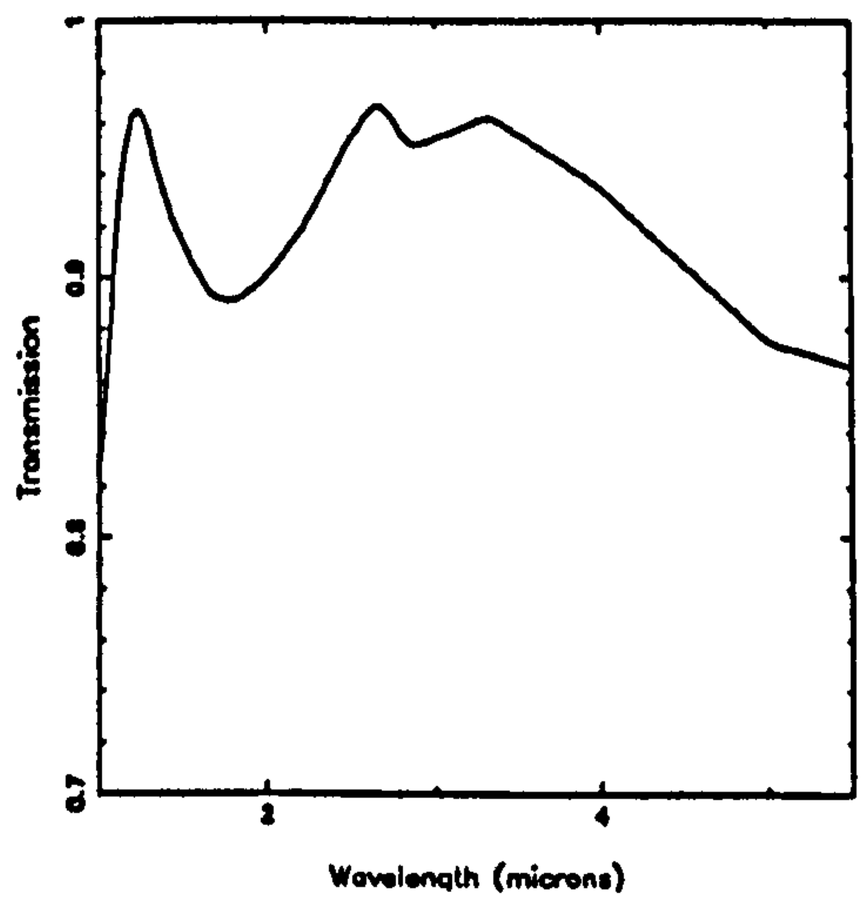


Figure 4.22: AR-coated ZnSe

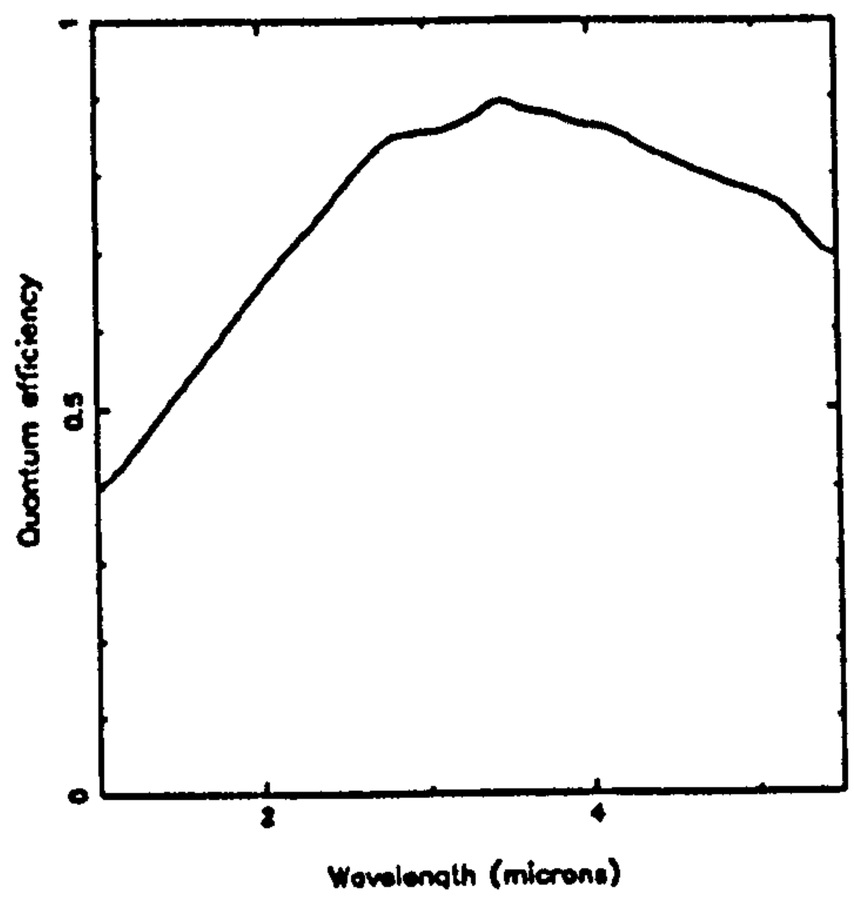


Figure 4.23: AR-coated InSb

In addition to the standard broad band filters, non-standard narrow band filters are also used. At the time of writing, the actual filter profiles were not available, and in order to use them in our simulations, we have modelled several such filter profiles. The most commonly used is the narrow band L (nbL) filter, which has a 1.8% FWHM at 3.6μm, and it is shown in figure 4.21. These model filter profiles are not expected to be very realistic, but should transmit the correct amount of flux to within 50% at least, and hence we have used them in the simulations that follow.

4.7.2 Lenses

The lenses are used to focus the collimated beam onto the detector. In the current IRCAM design, three basic plate scales are available — 0.6, 1.2, and 2.4$''$/pixel. The 0.6 and 2.4$''$/pixel scales are achieved with doublet lenses, whilst the intermediate plate scale is achieved with a singlet. All the lenses are made of zinc selenide (ZnSe), and are anti-reflection coated with a standard ITT AR coating, G4116/7. The transmission profile for a single AR-coated ZnSe lens is shown in figure 4.22, and the transmission profile for a doublet will be the square of this profile (Harris, private communication).

4.8 Detector

The detector is the final component. A full description of the SBRC 62 × 58 InSb + DRO array is given in chapter 3, and the detector model described there is used by SIR-

CAM in order to calculate the way that photo-electrons are processed. Firstly however, the detector material must convert the incoming astronomical and background photons into photo-electrons. The fraction of photons converted into photo-electrons is given by the quantum efficiency η. The raw semi-conductor quantum efficiency may be increased significantly with the use of a an anti-reflection coating, which reduces dielectric reflection losses Figure 4.23 shows the spectral dependence of the quantum efficiency for an AR-coated SBRC array at 77 K as measured at SBRC (Orias, private communication). Note that as yet SIRCAM does not deal with the temperature dependence of the quantum efficiency, discussed further in chapter 3.

4.9 Prediction of background and source fluxes

In this section we shall make some predictions of the detected photocurrents from both the sky background and a star of given brightness for IRCAM through each of the standard filters in the 0.6″/pixel mode. The corresponding background figures for the 1.2″ and 2.4″/pixel modes can be calculated by scaling proportionally to the pixel area.

We use the information gathered so far on atmospheric transmission, atmospheric emission, thermal emission, optical throughput, and so on. These predictions are made in the spectral mode of SIRCAM, and we begin by discussing the basic method.

4.9.1 Method

All the various optical transmission, reflectivity, and absorption spectra are input at a spectral resolution of 10^{-3} μm. Similarly the non-thermal sky background spectrum is input. The user chooses a particular atmospheric spectrum, defined by the water vapour content and airmass. Also chosen are the ambient temperature, the effective sky temperature, the pixel scale, and the optical dirt factors.

The non-thermal sky emission spectrum is convolved with the atmospheric transmission spectrum, and the atmospheric thermal emission is added, assuming Kirchoff's Law, by taking one minus the transmission to be the emissivity. This gives the background flux arriving above the telescope. The total warm optical throughput is calculated, and the sky emission spectrum is attenuated accordingly. The warm optical thermal emis-

sion is added, again by assuming that one minus the transmission equals the emissivity. This now gives the total background flux entering the cold instrument. This spectrum is then convolved with the cold throughput, defined by the lenses and detector quantum efficiency. We assume no flux is added by the cold optics.

We now have a spectrum of the unfiltered background flux in terms of photo-electrons per second per spectral bin per unit area per unit solid angle. Then, for each filter in turn, we convolve this spectrum with the cold filter profile, then summing across the profile to give us the normalised total detected background flux for the filter. Lastly, these numbers are multiplied by the telescope collecting area, pixel area, and fill factor, in order to scale the background to the telescope and instrument configuration in use.

At the same time, a scaled version of the Vega model spectrum is similarly convolved with the atmospheric transmission, the warm throughput, the cold throughput, and each filter profile, in order to calculate the signal detected through each filter for a star of known brightness per unit collecting area. Multiplying by the telescope area, this is scaled accordingly.

Thus the final results from this routine are the detected background flux in photo-electrons per second per pixel, and the total detected stellar flux in photo-electrons per second, through each filter.

4.9.2 Predicted backgrounds for the 0.6″/pixel mode

Obviously, the range of the free parameters is large, and here we shall present figures for one representative configuration only. We have predicted the background photocurrent in electrons per second per pixel that should be seen by IRCAM in the 0.6″/pixel imaging mode, on the UKIRT on Mauna Kea. The backgrounds for the other pixel scales can be calculated by scaling in proportion to the pixel area imaged on the sky.

We have used the following parameters :

- IRCAM on the 3.8 m UKIRT on Mauna Kea
- 0.6″/pixel image scale
- Warm optical temperature 275 K, effective sky temperature 253 K, and water

Filter	Signal (e^-/sec/0.6″ pixel)				
	1.00	1.25	1.50	1.75	2.00
J	1044	1280	1507	1727	1938
H	3935	4883	5813	6724	7614
K	8482	9111	9719	10305	10870
nbL	1.22×10^6	1.23×10^6	1.25×10^6	1.26×10^6	1.28×10^6
L′	2.85×10^7	2.94×10^7	3.03×10^7	3.10×10^7	3.17×10^7
M	1.43×10^8	1.48×10^8	1.52×10^8	1.56×10^8	1.59×10^8

Table 4.2: Predicted sky background for IRCAM on UKIRT vs. airmass

vapour content of 1.2 mm

- Nominal OH*, O_2, and infrared continuum non-thermal emission intensities
- Dirt factors for the primary, secondary, dichroic, and gold mirrors of 0.95, 1.00, 0.98, and 0.99 respectively

In table 4.2 we give the resulting background photocurrent per pixel as a function of airmass. Note that the backgrounds in the non-thermal (1–2 μm) regime change almost proportionally to the airmass, whereas the much larger thermal backgrounds seen between 3–5.5 μm do not change much, as they are dominated by emission from the telescope and warm optics. The background seen through the K filter is mainly thermal emission, with some non-thermal emission varying with the airmass.

It is important to note the backgrounds predicted for the thermal wavelengths. Assuming a full well depth of 10^6 e^-, the detector array will saturate in less 35 milliseconds at L′, and in about 7 milliseconds at M. As IRCAM currently uses a minimum read out time of 32 milliseconds, and given that these backgrounds do not allow for any slight baffling problems, it is apparent that imaging at L′ and M will be impossible in this observational configuration. Saturation may be prevented by using narrower band filters, such as the nbL filter, or by stopping down the present filters to reduce the throughput. Either solution implies a loss of signal to noise, but some imaging capability at these wavelengths would be better than none.

Also note the relatively low predicted background photocurrents for the non-thermal filters. At J, over five minutes on-chip integration is required to fill to half well, and thus

Filter	Signal (e^-/sec)				
	1.00	1.25	1.50	1.75	2.00
J	8.55×10^9	8.45×10^9	8.37×10^9	8.30×10^9	8.23×10^9
H	6.57×10^9	6.55×10^9	6.53×10^9	6.51×10^9	6.49×10^9
K	5.76×10^9	5.71×10^9	5.67×10^9	5.63×10^9	5.59×10^9
nbL	3.25×10^8	3.21×10^8	3.17×10^8	3.13×10^8	3.09×10^8
L′	2.95×10^9	2.88×10^9	2.81×10^9	2.75×10^9	2.69×10^9
M	7.53×10^8	7.08×10^8	6.69×10^8	6.36×10^8	6.07×10^8

Table 4.3: Predicted signal from $0^m.0$ star on UKIRT vs. airmass

become background limited. We can see that the K background is about double that at H, but because it is mainly thermal emission from the telescope, we would expect K to be far more stable than H, at which wavelength the background is due to the highly variable non-thermal emission processes.

4.9.3 Predicted stellar fluxes

Using the same model parameters as given above, we have also predicted the total number of photo-electrons per second detected from a $0^m.0$ star. Again, we have predicted this number for each of the standard filters, and again as a function of airmass. The results are shown in table 4.3

Note that through the filters in the cleaner windows (J, H, K), only small changes (5–10%) in detected flux are seen between 1 and 2 airmasses. This confirms the low value of extinction seen for these filters [83], although as mentioned earlier, we have neglected the rôle of molecular and aerosol scattering in our simulations, and values of extinction predicted from our modelling would underestimate that actually seen. At M, the change in detected flux is $\sim 20\%$, due mainly to the highly structured water vapour absorption seen through this filter, which increases with airmass.

4.10 Sensitivity predictions

We are now able to provide some astronomical sensitivity predictions, by drawing together the following components that we have derived or simulated in previous sections in this chapter, and from the detector modelling in chapter 3 :

- the incoming flux from a source of known brightness
- the total background flux, comprising atmospheric thermal and non-thermal emission, and thermal emission from the warm optics
- atmospheric and optical transmission
- atmospheric seeing profile
- detector characteristics, including quantum efficiency, dark current, read noise, fill factor, well depth, and spatial noise
- the model for the noise present after full data reduction

4.10.1 Parameters and methods

Combining all this information, we can derive a variety of sensitivity figures, as discussed below. However, before doing so, we shall list the most important parameters as follows :

- IRCAM on the 3.8 m UKIRT on Mauna Kea, with an ambient temperature of 275 K
- a telescope airmass of 1.5, an atmospheric water vapour content of 1.2 mm, a seeing FWHM of 1.5″, and nominal non-thermal atmospheric emission flux levels
- an AR-coated low doped SBRC 62×58 InSb + DRO array, with a dark current of $100\,e^-$/sec/pixel, a read noise of $375\,e^-$ RMS, and a mandatory read out at $\sim \frac{1}{2}$ well ($5 \times 10^5\,e^-$). An area fill factor of 0.9 is assumed.

The actual sensitivity calculations are made as follows. We 'reduce' the data according to the standard linear technique discussed at length in chapter 3 : the source and sky frames (each of equal on-chip integration and number of co-adds) both have a dark frame

(also same on-chip integration time and number of co-adds) subtracted, to remove the offset level and dark current. The source frame is then divided by the sky frame to flat-field it, followed by a rescaling and sky subtraction. In this case, the signal-to-noise (S/N) is calculated according to equation 3.96.

Beginning with the first source frame, the on-chip integration is continued until the desired S/N is reached for the given input flux. If this S/N is not reached before either (a) the detector is filled to $\frac{1}{2}$ well or (b) a maximum specified on-chip integration time is reached, the array is assumed to be read out, and a new on-chip integration commences. This continues until the desired S/N is finally reached. The maximum on-chip integration time is only reached at J and through the H_2 v=1–0 S(1) narrow band filter where the background is low — the maximum is specified as 300 seconds.

The final total integration time given includes the time taken to obtain all the source frames, and a corresponding number of sky frames with equal on-chip integration time. Not included is the time taken to obtain the a similar set of dark current frames. The sky frames are included as the best noise performance is achieved only when using quasi-simultaneous sky frames as the flat-field, and using 'off-line' flat-field exposures of the dome for example has not yet been proven as effective. The dark current frames can be obtained during dusk or dawn, as the dark current is low, and these frames are really measuring the stable bias offset.

Whilst these assumptions are reasonably realistic, there is clearly some leeway for improvement in reducing residual noise, by median stacking of the sky frames, or by obtaining many more dark current exposures off-line, for example. Nevertheless, the figures derived with this technique provide useful indications of the sensitivity of IRCAM on the UKIRT, with the following precautions :

- It must be stressed that the sensitivity figures presented here are essentially detection limits, as opposed to measures of photometric accuracy. That is, a 20σ detection here implies that a source can be seen at a level twenty times the noise. The conventional infrared wisdom would be to invert this number and say that the source brightness has been defined to an accuracy of 1/20th or 5%. Although this method is roughly valid for an AC-coupled photometer, it cannot be applied to infrared imaging, where the accuracy to which a source's brightness is known is

defined considerably differently, dependent upon the accuracy of the sky measurement, the pixel scale, the fill factor, and the reduction technique. This comment is particularly relevant to point sources, and is discussed in more detail below.

- Although the non-linearity of the SBRC array is of importance when attempting accurate relative calibration, it can be neglected in this context, as it only implies an effective attenuation of the brighter sources by $\sim 5\%$, thus not seriously affecting sensitivity predictions.

- The examples given of extended source imaging in emission lines *do not* include the time taken to image the same source at a nearby continuum wavelength, required in order to remove the underlying continuum emission. This process will effectively double the time taken to image emission line sources with any appreciable continuum.

- Finally, even though these predictions are as realistic as possible, they still represent an ideal case, and the presence of $1/f$ noise, residual structure in the flat-fielded frame, or low level extended nebulosity in the source frames, may degrade the actual astronomical results to some extent.

4.10.2 Extended sources

It is relatively easy to determine what is meant by the sensitivity of a camera to an extended source of a given surface brightness. We shall assume it means the level to which a pixel containing the given source brightness rises above the local noise. We have calculated the total integration time taken to reach both 3 and 20σ (S/N = 3 and 20) for a range of surface brightnesses, for each of the standard near infrared broad band filters, as well the nbL narrow band filter at $3.6\,\mu$m, and two representative emission line filters, namely the H_2 v=1–0 S(1) filter at $2.122\,\mu$m in the non-thermal, and Brα at $4.05\,\mu$m in the thermal infrared. These figures are only shown for the $0.6''$ and $2.4''$/pixel modes, as the corresponding figures for the $1.2''$/pixel mode may be calculated by suitable interpolation. The results are shown in figures 4.24 to 4.39, and also shown on the figures is the detected sky background, and the longest on-chip integration time used in each case, virtually always determined by the time taken to fill to $\frac{1}{2}$ well on the sky background.

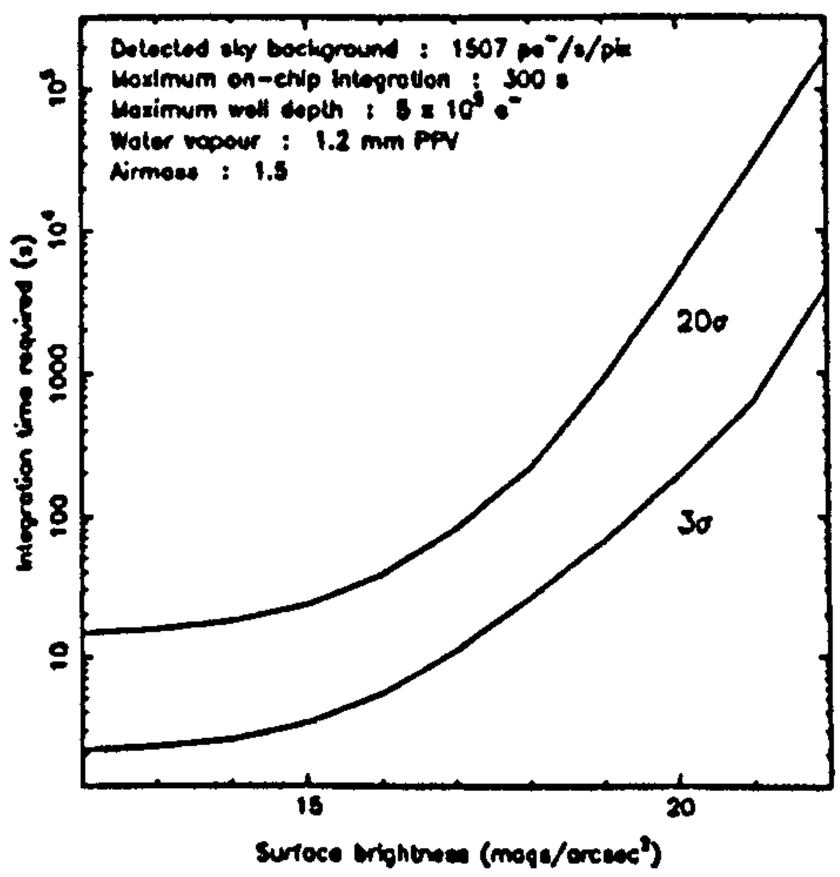


Figure 4.24: J : extended : 0.6″/pix

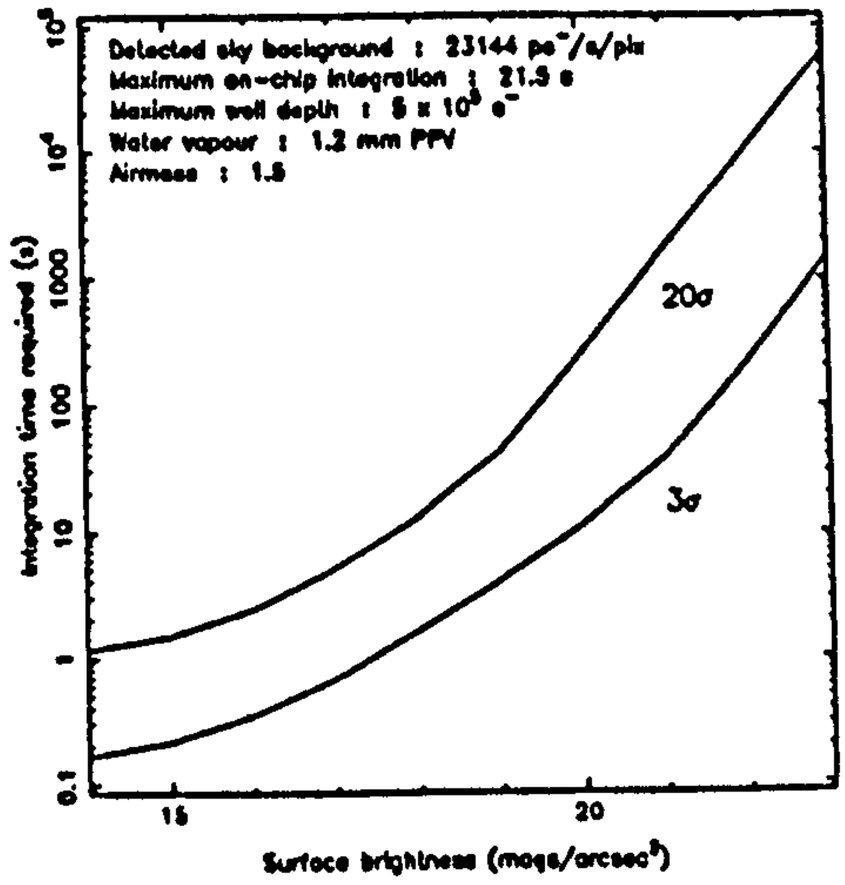


Figure 4.25: J : extended : 2.4″/pix

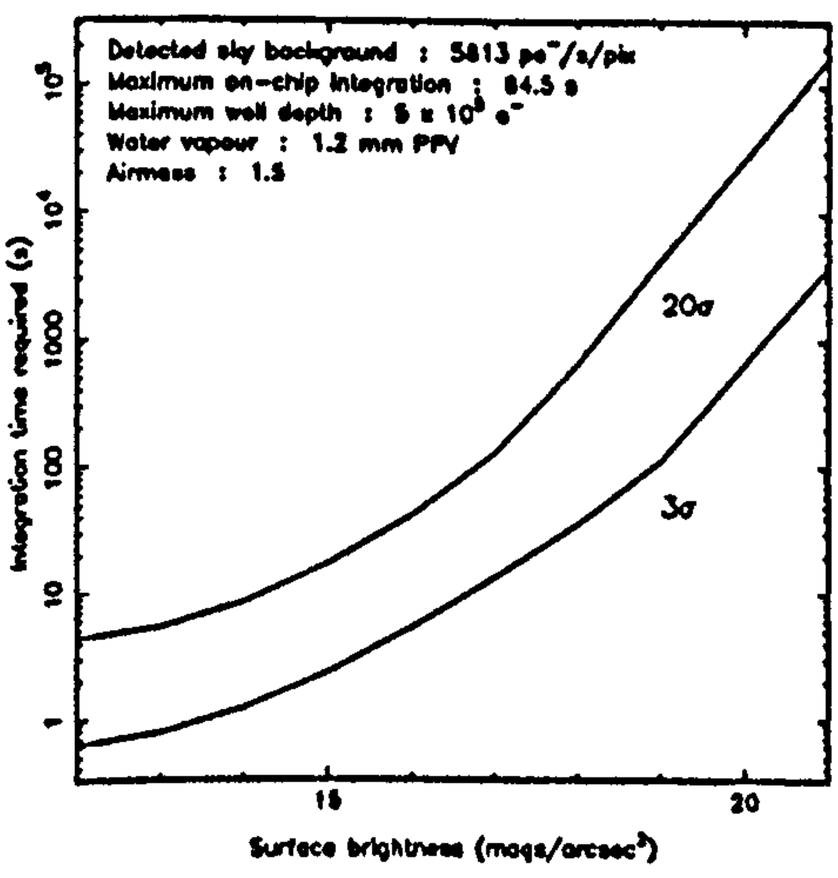


Figure 4.26: H : extended : 0.6″/pix

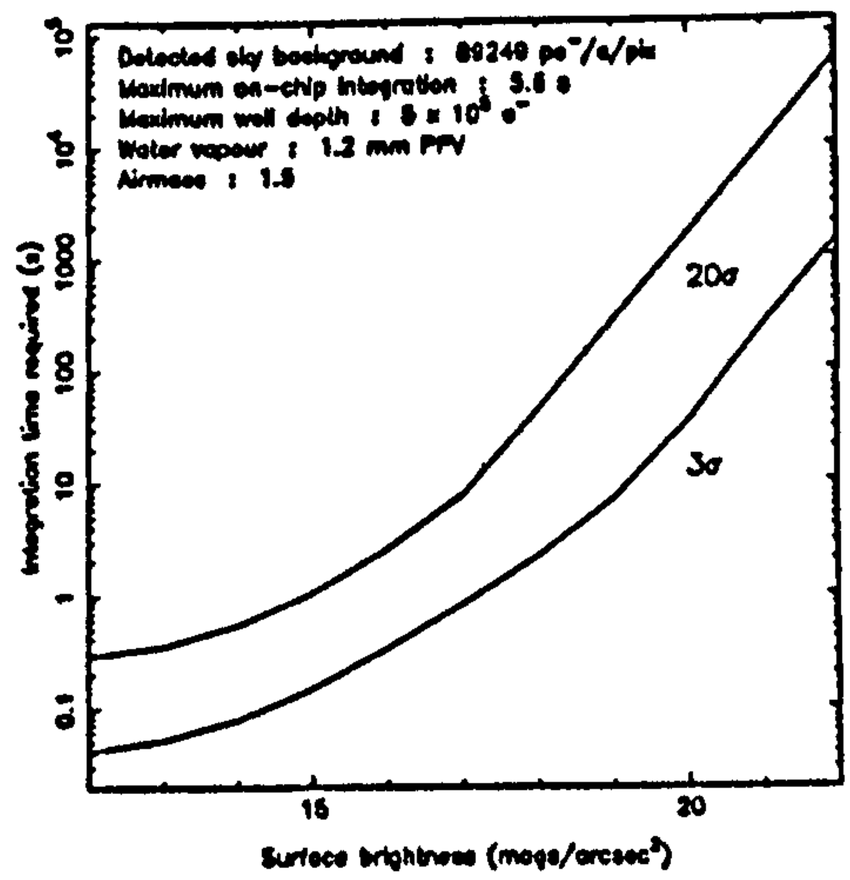


Figure 4.27: H : extended : 2.4″/pix

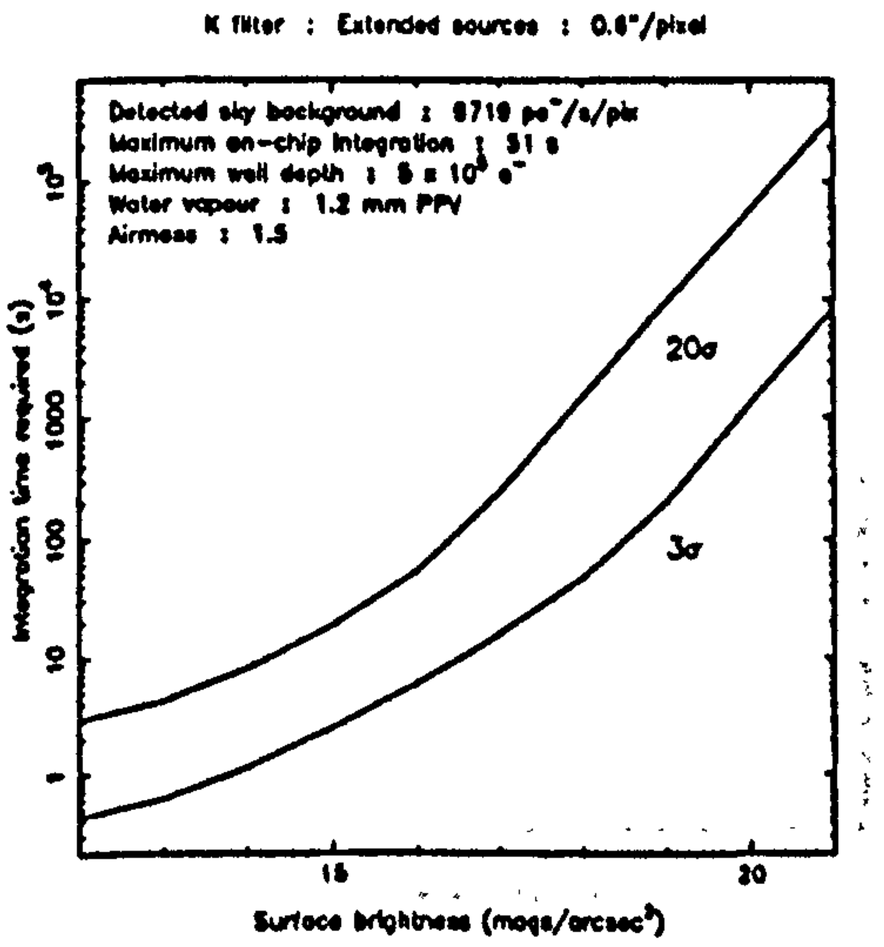


Figure 4.28: K : extended : 0.6″/pix

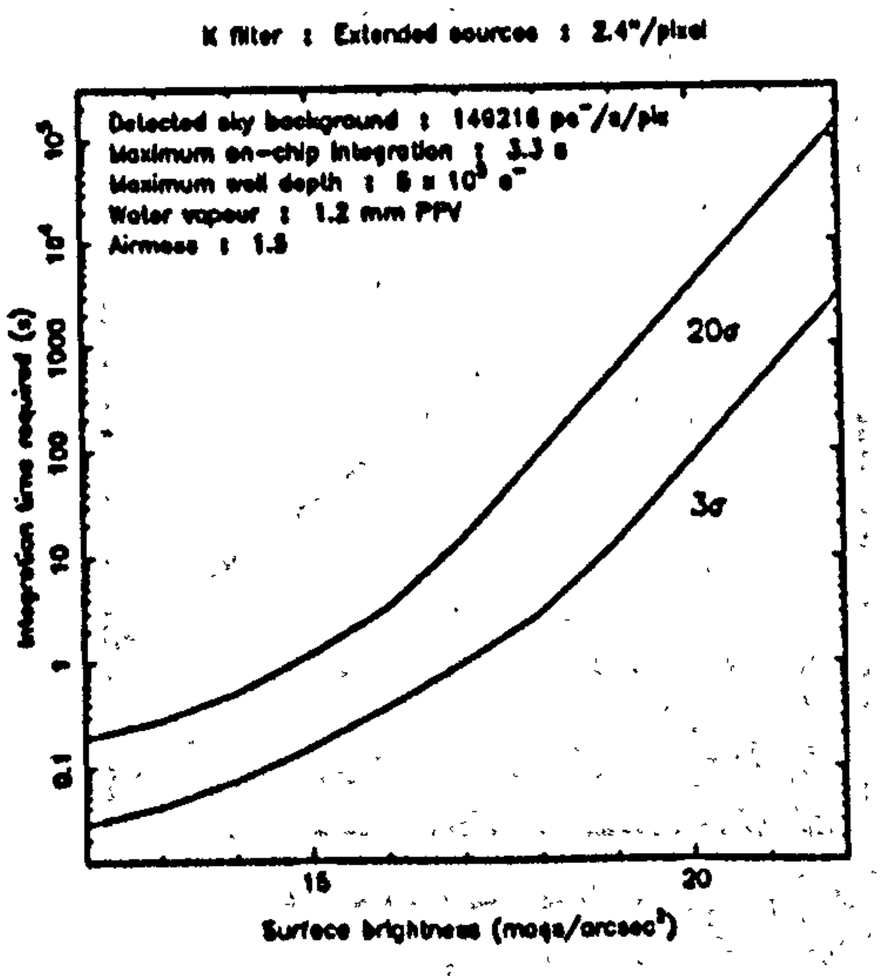


Figure 4.29: K : extended : 2.4″/pix

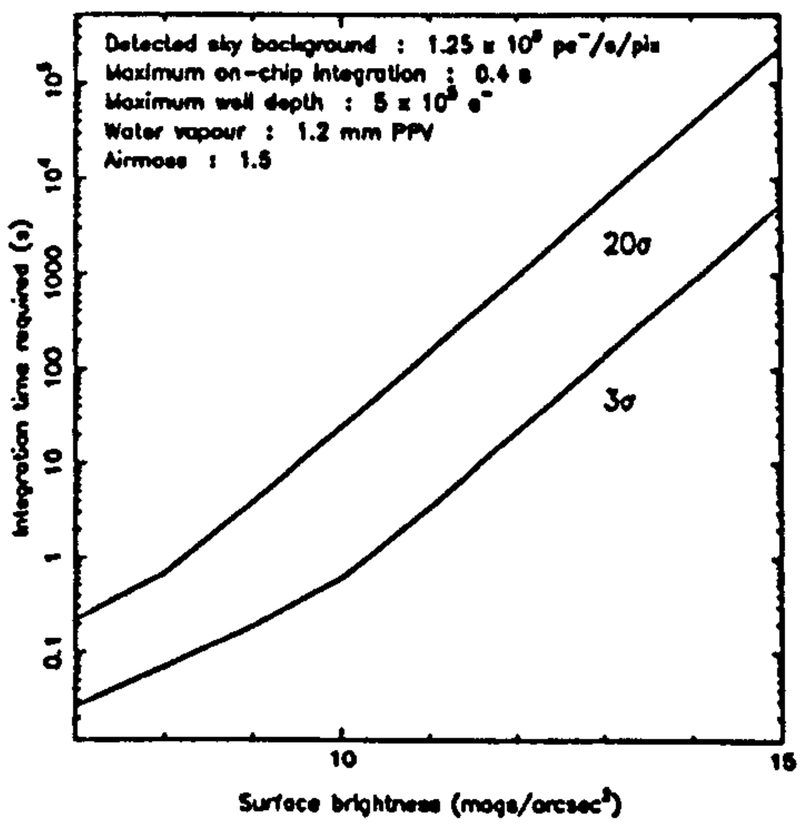


Figure 4.30: nbL : extended : 0.6″/pix

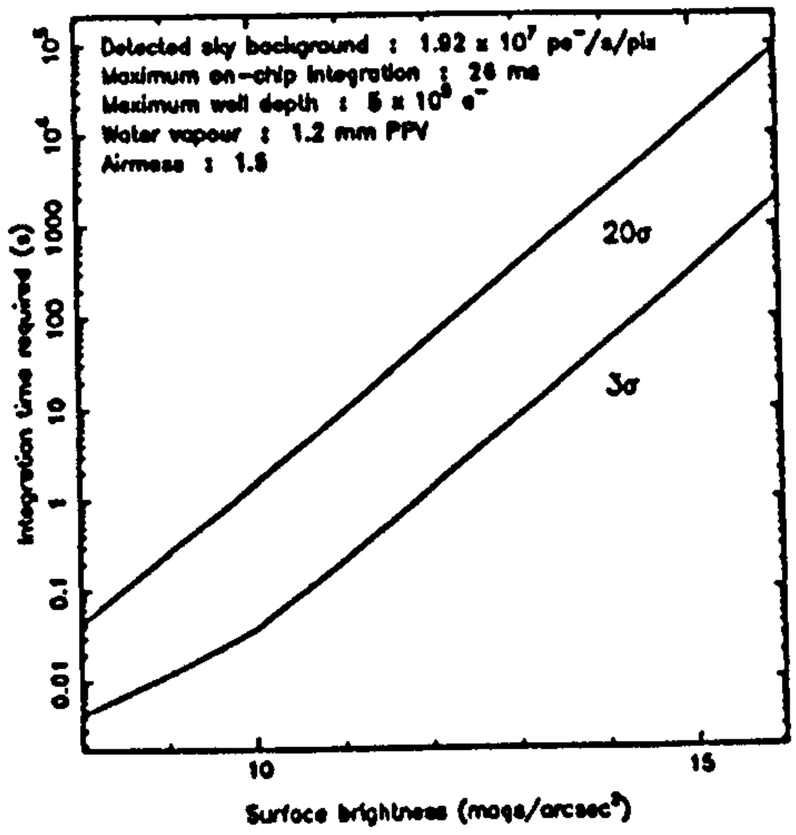


Figure 4.31: nbL : extended : 2.4″/pix

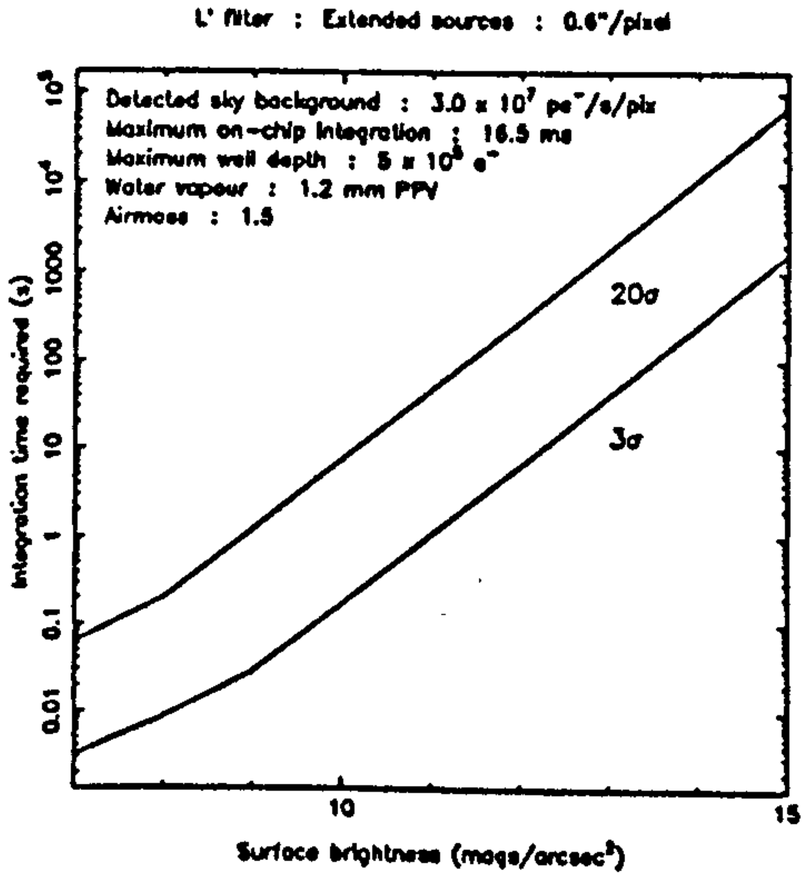


Figure 4.32: L′ : extended : 0.6″/pix

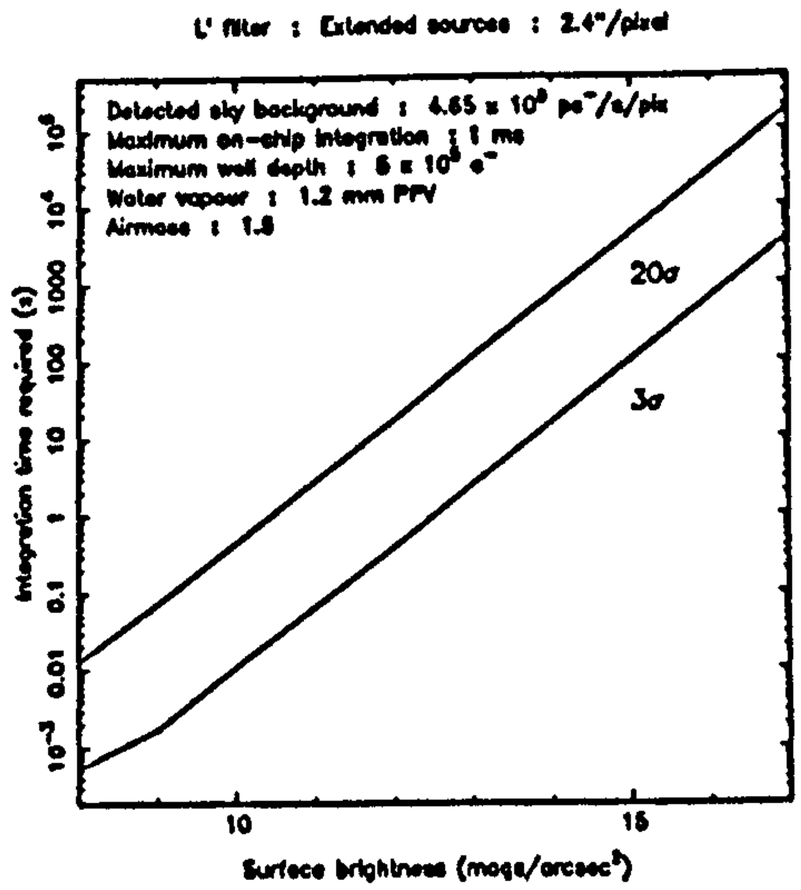


Figure 4.33: L′ : extended : 2.4″/pix

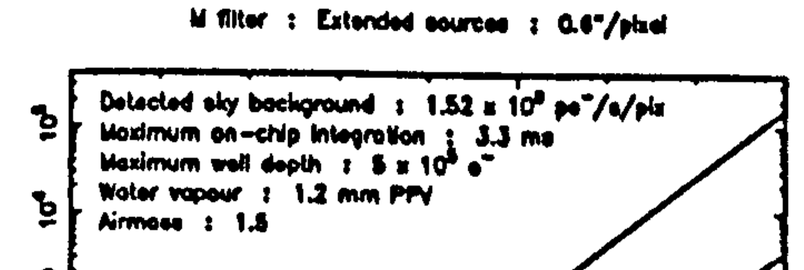


Figure 4.34: M : extended : 0.6″/pix

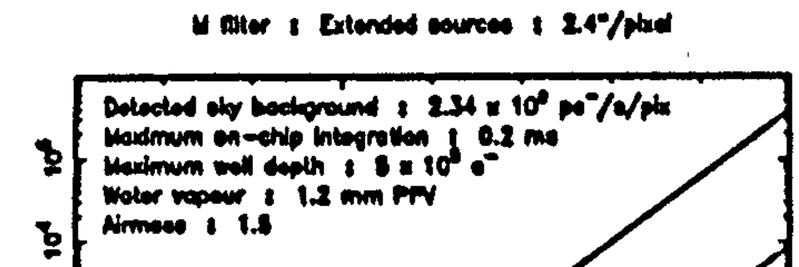


Figure 4.35: M : extended : 2.4″/pix

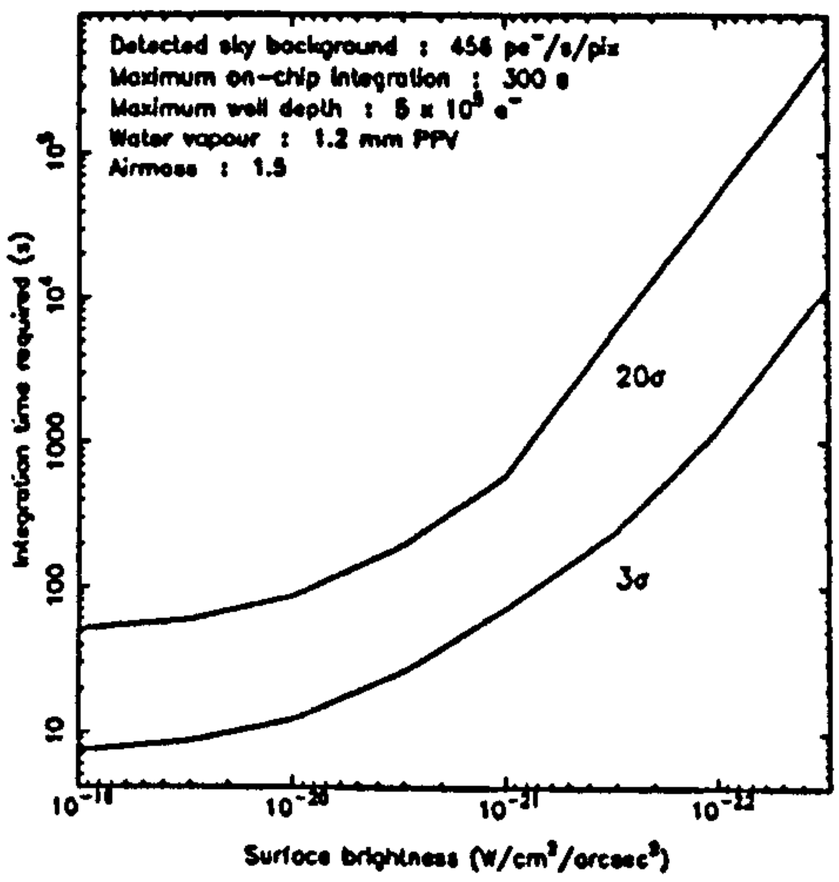


Figure 4.36: S(1) : extended : 0.6″/pix

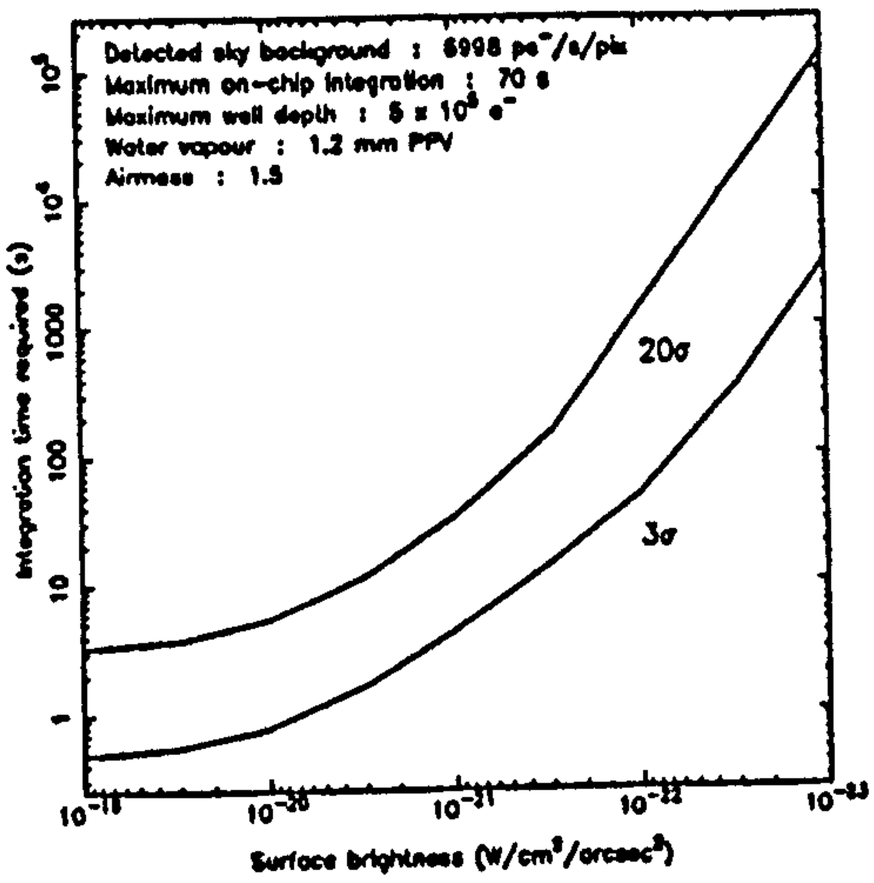


Figure 4.37: S(1) : extended : 2.4″/pix

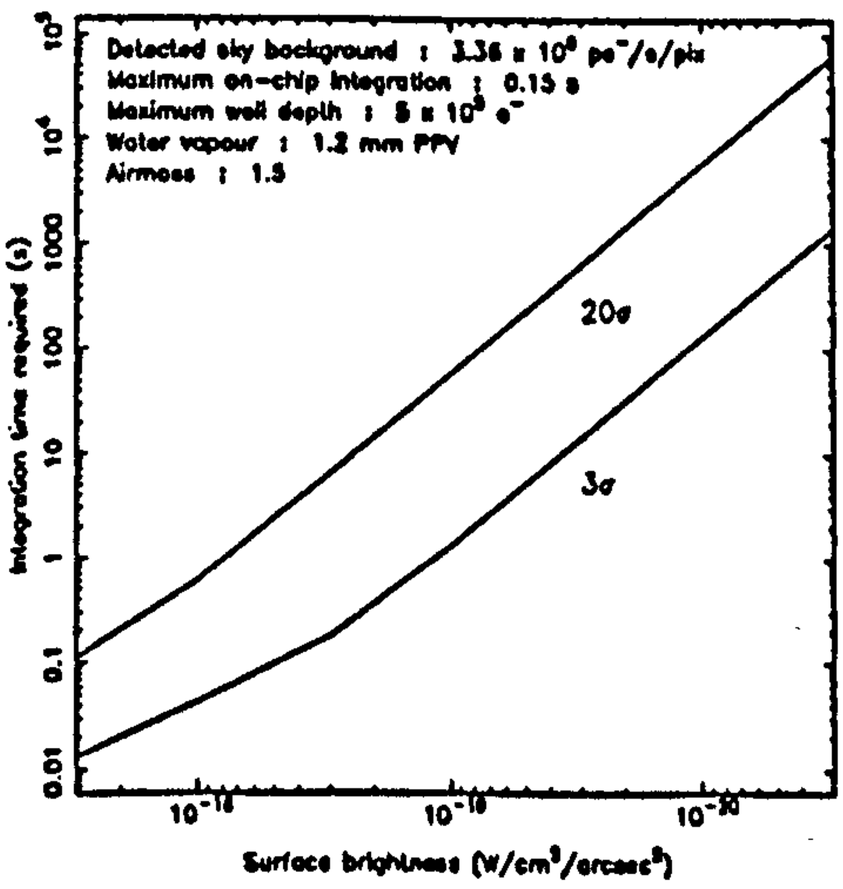


Figure 4.38: Brα : extended : 0.6″/pix

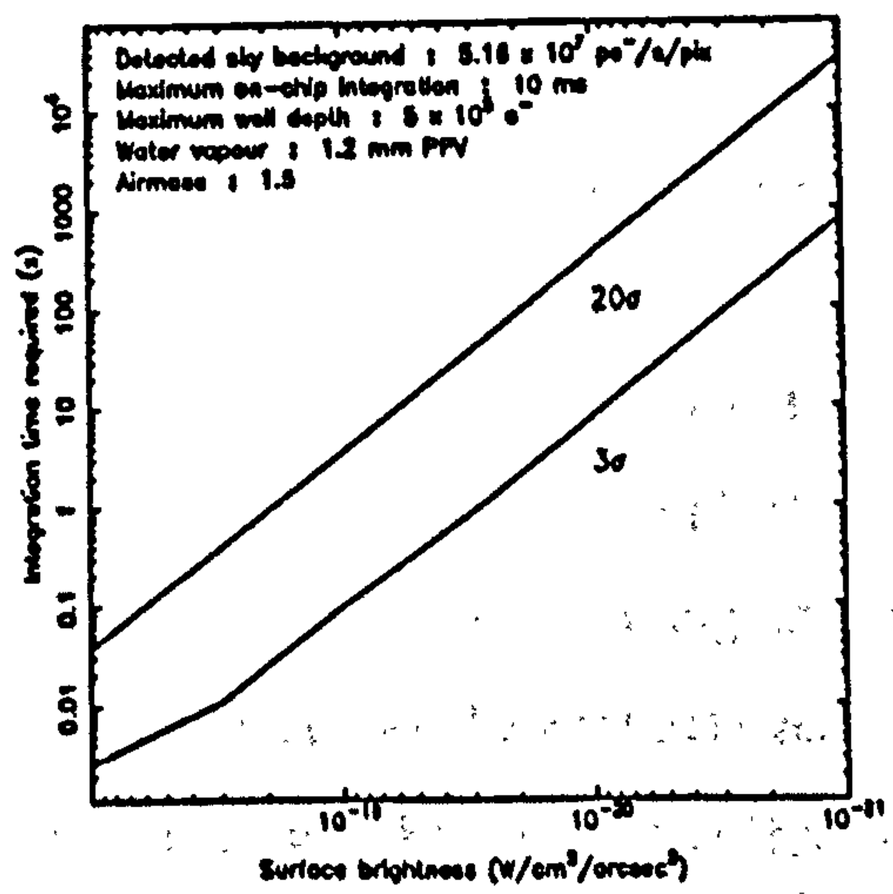


Figure 4.39: Brα : extended : 2.4″/pix

4.10.2.1 Discussion

We shall now briefly discuss the results. Firstly for the non-thermal wavelengths, J, H, K, and the H_2 v=1–0 S(1) line. There are three sections to each curve, starting with the section for the brighter fluxes, where the time taken to reach a given signal-to-noise asymptotically tends towards a constant value, regardless of the flux. As discussed in section 3.8, this is because the S/N for the brighter sources is determined by the accuracy of the the flat field, and not the source shot noise. On the intermediate section of the curves, we are read noise limited, and the S/N for a given source improves proportionally with time. Put another way, the faintness limit seen at a fixed sensitivity decreases with time. The third section of the curves is the background limited regime, where the S/N improves as the square root of time, or in these diagrams, the faintest surface brightness seen at a given S/N improves as the square root of time. The slight inflections in the curves between the read noise and background noise limited regimes occurs at the point when we change from a single on-chip integration to multiple co-adds, each of the latter being background limited in general.

The thermal filters (nbL, L′, M, Brα) show a small read noise limited section at the very short integration times, but rapidly move into the background limited regime, improving as the square root of time. It is worth noting at this point that of the thermal imaging configurations shown, many are actually impractical for the current IRCAM due to the excessively high read rates required — these are nbL 2.4″/pixel, L′ 0.6″ and 2.4″/pixel, M 0.6″ and 2.4″/pixel, and Brα 2.4″/pixel. Clearly, imaging through the current broad band L′ and M filters is impossible at any of the available scale sizes, and for true thermal imaging to be possible, steps must be taken to reduce the thermal background, either by stopping down the current filters, or using narrow band filters instead. However, as the nbL and Brα are background limited at 0.6″/pixel, binning these modes in software would reach the corresponding sensitivity limits predicted for the 2.4″/pixel mode, the only disadvantage being that the total field of view in the 0.6″/pixel mode is substantially less than in the wide field mode.

4.10.2.2 Comparisons

In order to assess the kinds of observations that should be possible if these sensitivity predictions are borne out in practice, we will now examine some typical surface brightnesses for a range of sources.

Imaging of nearby normal spiral galaxies in the near infrared (J, H, K) is of interest as these wavelengths trace the distribution of late type giants, which delineate the underlying mass distribution of these galaxies, in contrast to optical (B, V) light, which trace the younger O and B stars predominantly located in the star forming spiral arms [133]. For M31 for example, at a distance of 20 Kpc out along the major axis, the K surface brightness has fallen to $\sim 19^{m}.0/\square''$, with a H – K of about $0^{m}.0$ [17]. At K in the $0.6\square''$/pixel mode, IRCAM should be able to make a 3σ detection to this level in $\sim$ 1000 seconds, whilst a 20σ detection would be possible in $\sim$ 2000 seconds in the $2.4\square''$/pixel mode. With a typical J – K of $\sim 1^{m}.0$ [133], the same sensitivity limits could be reached for a J surface brightness of $20^{m}.0/\square''^{2}$ in about half those times. Similar surface brightnesses have been measured at similar distances from the nuclei of several nearby spirals (e.g. NGC7331, NGC4565, NGC4216, NGC7814) [133], indicating that high spatial resolution infrared imaging for these galaxies is now possible.

In a more extreme example, it is thought that a detection limit of $K \sim 20^{m}.0$ is likely to give a reasonable chance of detecting primæval galaxies, for a wide range of evolutionary scenarios, or will at least help constrain galaxy evolution models in the event of a significant null detection [87]. Assuming an integrated magnitude of $K = 20^{m}.0$ evenly distributed over 5 square arcseconds ($\sim 2.5\square''$ diameter), we obtain a surface brightness of $\sim 21^{m}.75$ at K. Clearly, from figure 4.28, we can see that even a 3σ detection will take a very long time with IRCAM in the $0.6\square''$/pixel mode. A 3σ detection may be possible however in about 30 minutes in the $2.4\square''$/pixel mode. In the latter case however, much of the flux will fall in a very few pixels, making it hard to distinguish from random bad pixels and cosmic ray events. Some slight gains may be obtained by improved dark subtraction and flat fielding, but probably not significantly, as we are basically background limited already. Note however that primæval galaxy models are relatively uncertain, and integrated fluxes as bright as $K \sim 18^{m}.0$ are possible (Yates, private communication), bringing them well within the detection limits predicted here.

Extended emission in the H_2 v=1–0 S(1) line at 2.122μm is seen in many galactic star forming regions [20], and as well as spectroscopy, high spatial resolution imaging in this line will be important in determining the exact geometrical relationships between outflows, shocks, ionisation fronts, and excitation sources. The Orion Nebula is the brightest known source in the H_2 S(1) line, and is well studied at this wavelength [21]. With a typical surface brightness of 10^{-20} W/cm^2/$\square''$, it should be easily observable with IRCAM, and indeed this experiment has already been performed successfully. More typical and challenging is the H_2 S(1) emission associated with the highly luminous outflow in the DR21 star forming region [46]. This emission has been observed in a large beam to a 3σ outer contour of 5×10^{-23} W/cm^2/$\square''$. A similar 3σ detection at this level should be possible with IRCAM in $\sim$ 200 seconds in the 2.4$''$/pixel mode, or $\sim$ 1 hour in the 0.6$''$/pixel mode.

Near infrared H_2 emission has also been detected in extragalactic systems, including the Seyfert galaxy NGC1068 [143] and luminous starburst galaxies such as NGC6240 and Arp220 [73]. In NGC1068, a mean surface brightness of 4.4×10^{-22} W/cm^2/$\square''$ was observed in an 8$''$ diameter aperture, whilst mean surface brightnesses of 7.5×10^{-22} and 1.25×10^{-22} W/cm^2/$\square''$ were observed in 5$''$ diameter apertures on the nuclei of NGC6240 and Arp220 respectively. If these do in fact represent the mean surface brightnesses, then we could expect to obtain a signal-to-noise of 20 in 1000 seconds and 1 hour on NGC6240 and NGC1068 in the 0.6$''$/pixel mode, whilst the fainter Arp220 would reach an S/N of 20 in about 1000 seconds in the 2.4$''$/pixel mode. Of course, it is likely that the sources will be structured within the 8$''$ and 5$''$ beam sizes. In this case, the emission knots would be detected more quickly.

Finally, the Brα line of ionised hydrogen at 4.05μm has a typical mean surface brightness of $\sim 5 \times 10^{-21}$ W/cm^2/$\square''$ over a 5$''$ diameter aperture in many galactic HII regions (Geballe, private communication), and IRCAM should be able to achieve a 3σ detection at this level in $\sim$ 1000 seconds in the 0.6$''$/pixel mode. The galactic centre has been imaged in the Brα line by the University of Rochester group [39], clearly indicating that such experiments are possible with infrared array cameras. Their outer contour level of $\sim 4 \times 10^{-20}$ W/cm^2/$\square''$ should be detectable at the 20σ level in $\sim$ 600 seconds by IRCAM in the 0.6$''$/pixel mode. The Rochester group made observations at ten positions in the galactic centre, observing at two nearby continuum wavelengths also, all within one hour. These observations were made with a pixel scale of 0.42$''$/pixel on

the 3 m IRTF, and although a direct comparison between their data and our predictions is not possible due to insufficient information in their paper, the IRCAM prediction is probably correct to within a factor of two.

4.10.3 Point sources

It is considerably harder to define a predicted sensitivity to point sources for an imaging system. Whether a point source is seen or not depends on the seeing and pixel scale, as this will affect the spread of the source photons over the array. When discussing a 'detection' sensitivity, should the signal be defined as the integrated flux from the whole star, and if so, over what number of pixels should this signal be calculated from? Is the noise is defined as the RMS variation in the sky, over which the source must be 'seen', or is it the noise on some measurement of the sky which is subtracted from the source? After some experimentation with various ways of calculating a suitable 'figure of merit', we have decided to use the following definition of point source detection. We define the point source sensitivity much as for the extended sources discussed above, in that we calculate the degree to which a given pixel is detectable above the local noise. However, rather than the simple case where the source flux in the pixel of interest is simply defined by the surface brightness in magnitudes per square arcsecond (for example), we define the source flux as that contained in the brightest pixel in a typical seeing profile. For example, with 0.6″/ pixels and a seeing FWHM of 1″, approximately 20% of the total stellar flux falls within the bounds of the brightest pixel, with the remaining 80% distributed in the other, fainter pixels. In the current definition, it is the peak pixel that we are trying to detect above the local noise.

This approach is not without problems however, which must be noted. For well sampled seeing profiles, the peak pixel contains much the same fraction of the total stellar flux regardless of precisely where on the pixel the profile centroid lies. In undersampled modes, this is not the case. For example, in the 2.4″/pixel mode, the peak pixel brightness may vary by a factor of ~ 4, depending on whether the profile centroid lies at the centre of a pixel, or at the intersection of four pixels. The question of incomplete area fill factor worsens this situation.

Also, it must be noted that a detection of a point source in the 2.4″/pixel mode must

Scale size	Centre	Intersection	Mean
0.6″/pixel	0.239	0.151	0.195
1.2″/pixel	0.632	0.212	0.422
2.4″/pixel	0.950	0.198	0.574

Table 4.4: Fraction of total point source flux contained in brightest pixel

regarded with extreme suspicion, if the detection is based on one one imaging position only. The flux from a point source in this mode will often fall almost completely within one pixel, but with almost 4000 pixels in the array, we also expect statistically to see about 10 single pixel 3σ deviations from the mean sky level, and cosmic ray events could easily simulate higher signal to noise detections. A suspected detection of this kind must be confirmed by moving the array on the sky and seeing if a source again appears in the same place relative to fixed features on the sky.

In order to provide some representative sensitivity predictions, we have simulated Lorentzian seeing profiles with a FWHM of 1.5″, placing the profile centroid at the centre of a pixel, and then at the intersection of four pixels, assuming an area fill factor of 0.9 throughout. Taking the mean of the two peak pixel brightnesses, we define the mean fraction of the total stellar flux that falls within the brightest pixel, and multiplying the incoming stellar flux by this fraction gives us the flux we then try to detect above the local noise. This fraction was calculated for each of the three scale sizes, and the results are presented in table 4.4.

It is important to note that although using this mean peak pixel fraction parameter will not lead to any unreasonable errors for any position of the profile centroid in the 0.6″/pixel mode, it may well be seriously in error for the undersampled scale sizes, in particular the 2.4″/pixel mode, where the peak pixel may be almost two times greater or three times less than the mean, depending on exactly where the star lands on the array. However, some indication of the variation in sensitivity is given by the ranges listed in table 4.4, and this should be instructive to anyone surveying for point like sources in the 1.2″ and 2.4″/pixel modes.

From what we have discussed, it is clear that sensible point source sensitivities in the 2.4″/pixel mode are very difficult to define, and we do not present figures for this mode. Some indication of the quality of 2.4″/pixel imaging observations of point sources may

be obtained by examining the simulated images later in this chapter.

We now look at the sensitivity figures derived under these assumptions. We show the point source sensitivity for each of the near infrared filters (J, H, K, L′, M) and the narrow band L filter also, in the higher resolution scale sizes, 0.6″ and 1.2″/pixel. The results are shown in figures 4.40 to 4.51.

It is important to stress again that the sensitivity figures shown here are only detection limits, and do not represent photometric accuracy. The photometric accuracy actually achieved for point sources will rely heavily on the reduction techniques used, depending on whether aperture of profile fitting photometry is used, for example, and on the accuracy to which the source and sky fluxes can be determined in the presence of residual flat field structures, ghosting, and non-linearity. The area fill factor will also be important in determining photometric accuracy, particularly in the undersampled modes [84]. We shall explore this problem further below.

4.10.3.1 Discussion

The origin of the various sections of the point source sensitivity curves presented is as discussed for the extended sources above. Note the one small difference for J in the 0.6″ and 1.2″/pixel modes : for very bright stars, the sensitivity curve actually turns up, implying that the brighter the source the *longer* it takes to detect it to a given level. Whilst this might appear strange at first, the reason becomes apparent on examination. When the source is much brighter than the background, it will determine the maximum on-chip integration time, if saturation is to be avoided. As the source grows brighter, this on-chip integration time grows shorter, and the amount of sky background flux collected in the corresponding sky frame grows smaller. The level in the sky frame becomes less well determined, and as it is used as the flat field, the overall noise increases, reducing the signal-to-noise for the bright source. Clearly though, this presents no fundamental limit to the measurement of bright sources, as the corresponding on-chip integration times are short, and more sky frames may be obtained to improve the flat fielding, without adding serious overheads.

The figures presented here appear to suggest that the 1.2″/pixel mode is just slightly more sensitive to faint point sources than the 0.6″/pixel mode. This is because the mean

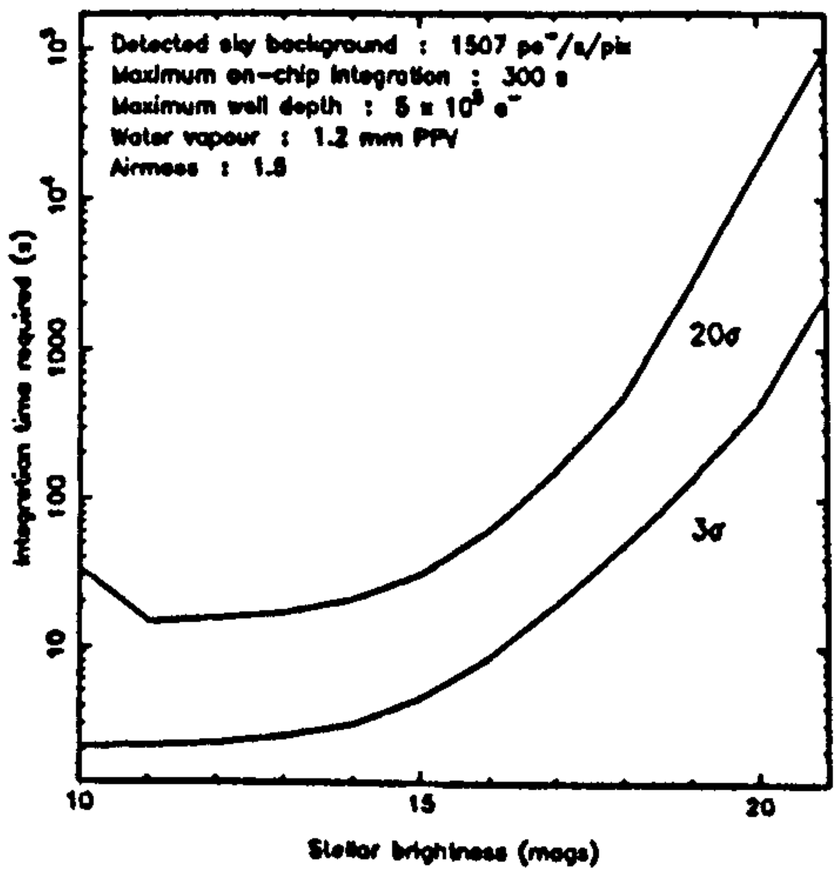


Figure 4.40: J : point : 0.6″/pix

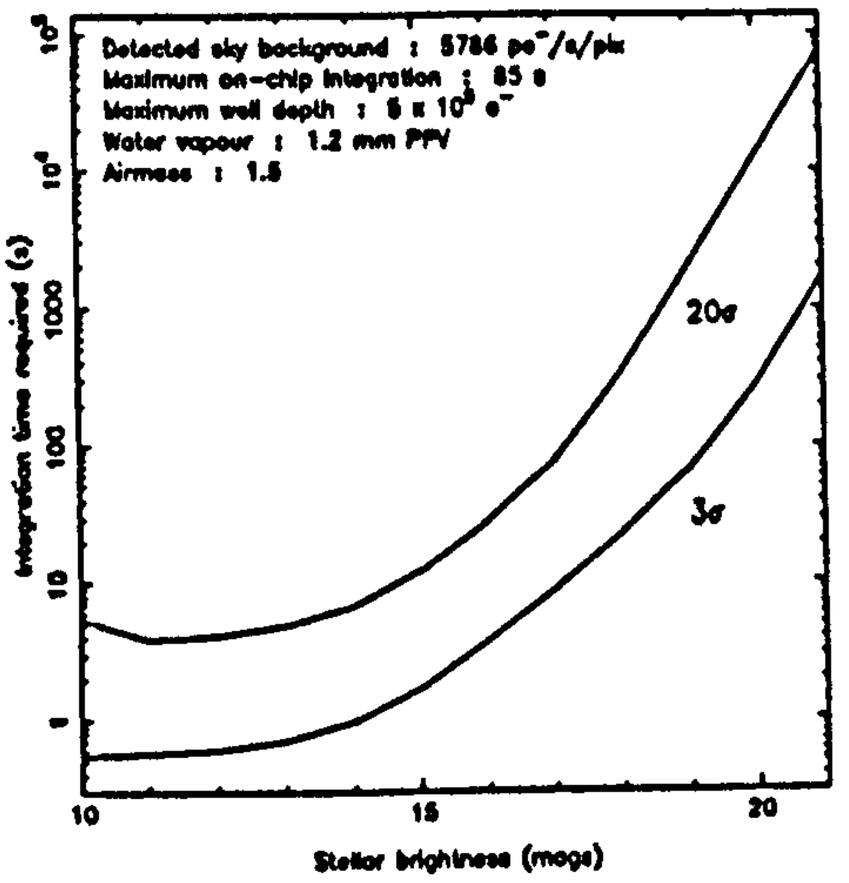


Figure 4.41: J : point : 1.2″/pix

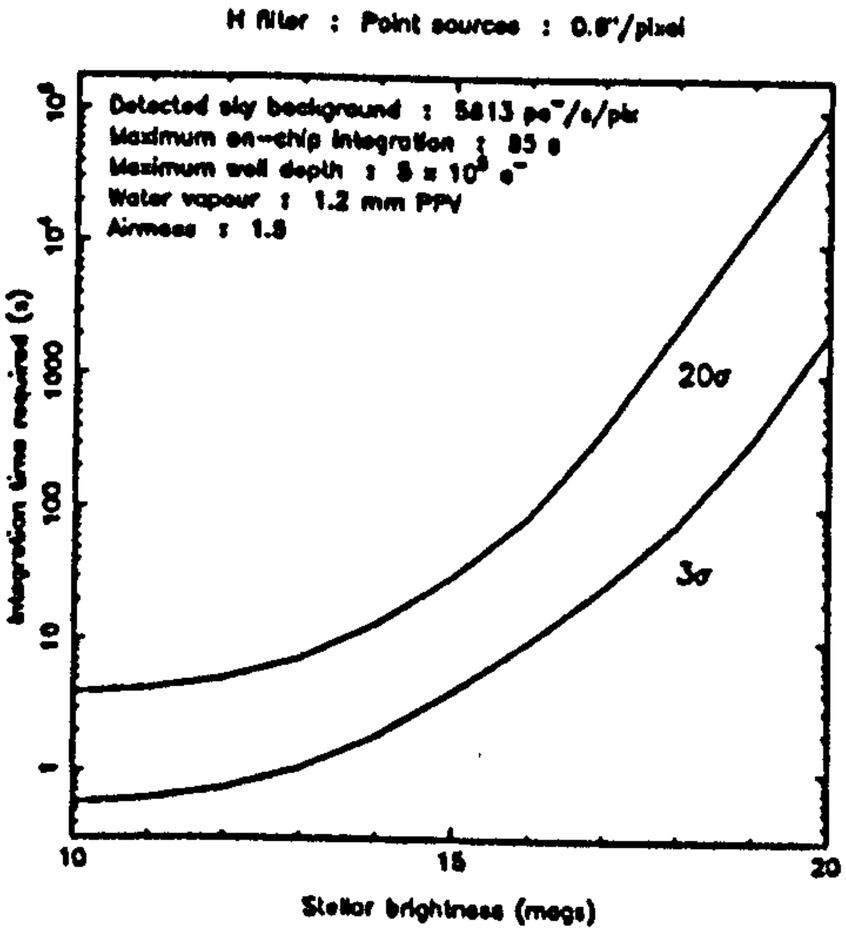


Figure 4.42: H : point : 0.6″/pix

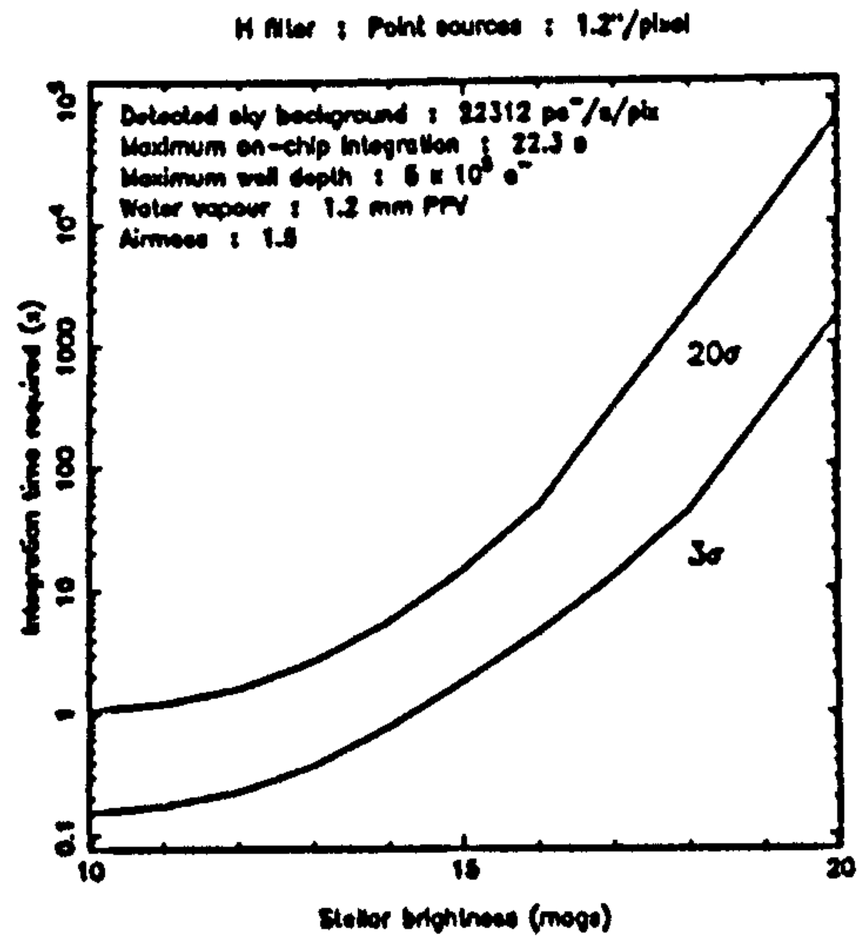


Figure 4.43: H : point : 1.2″/pix

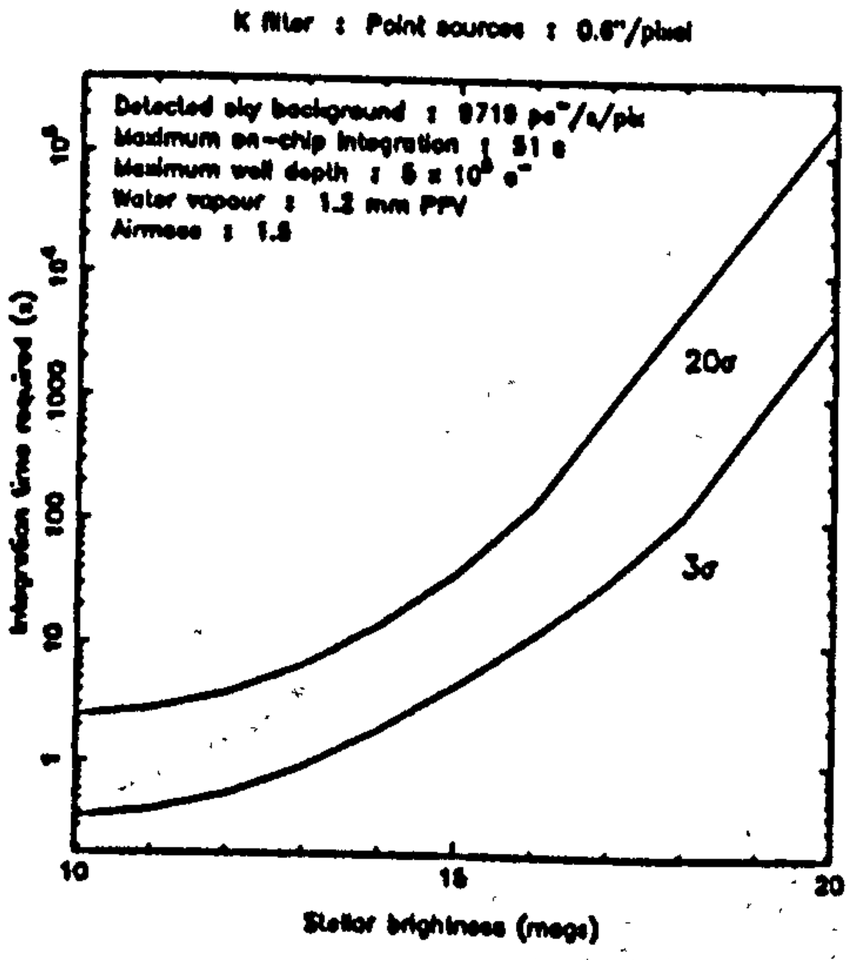


Figure 4.44: K : point : 0.6″/pix

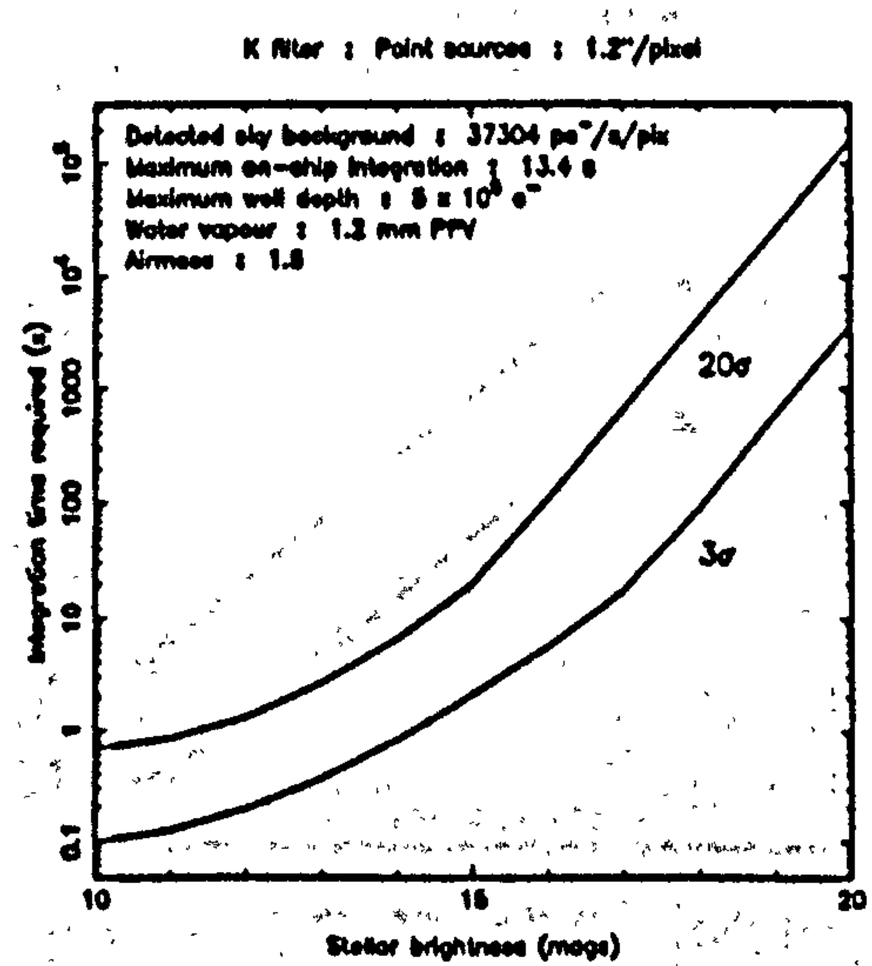


Figure 4.45: K : point : 1.2″/pix

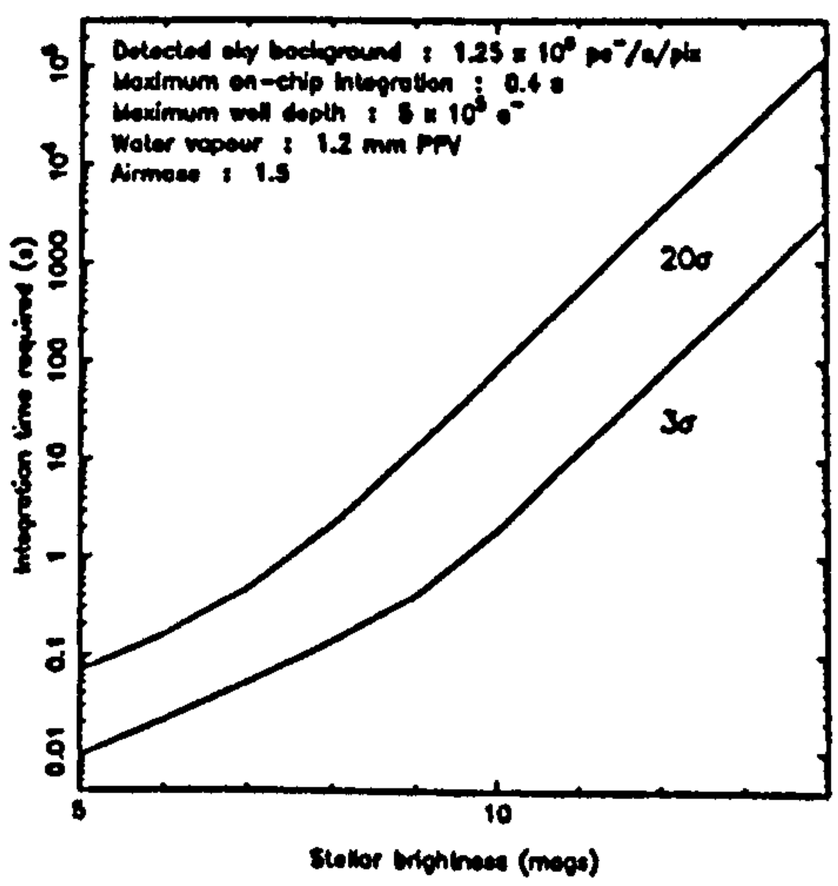


Figure 4.46: nbL : point : 0.6″/pix

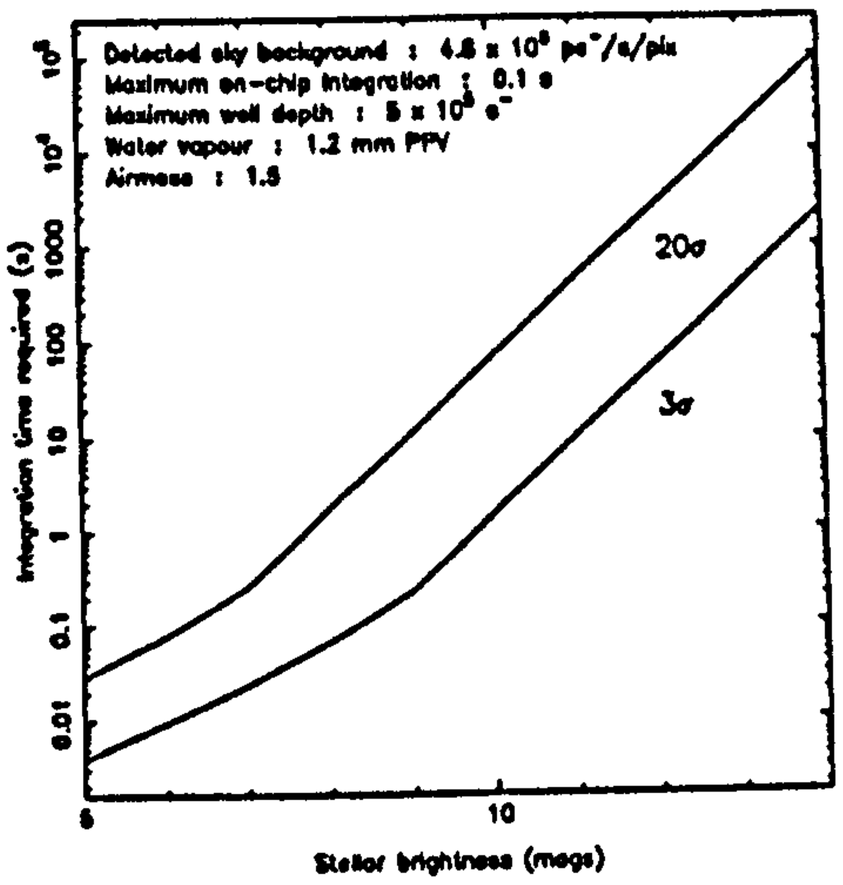


Figure 4.47: nbL : point : 1.2″/pix

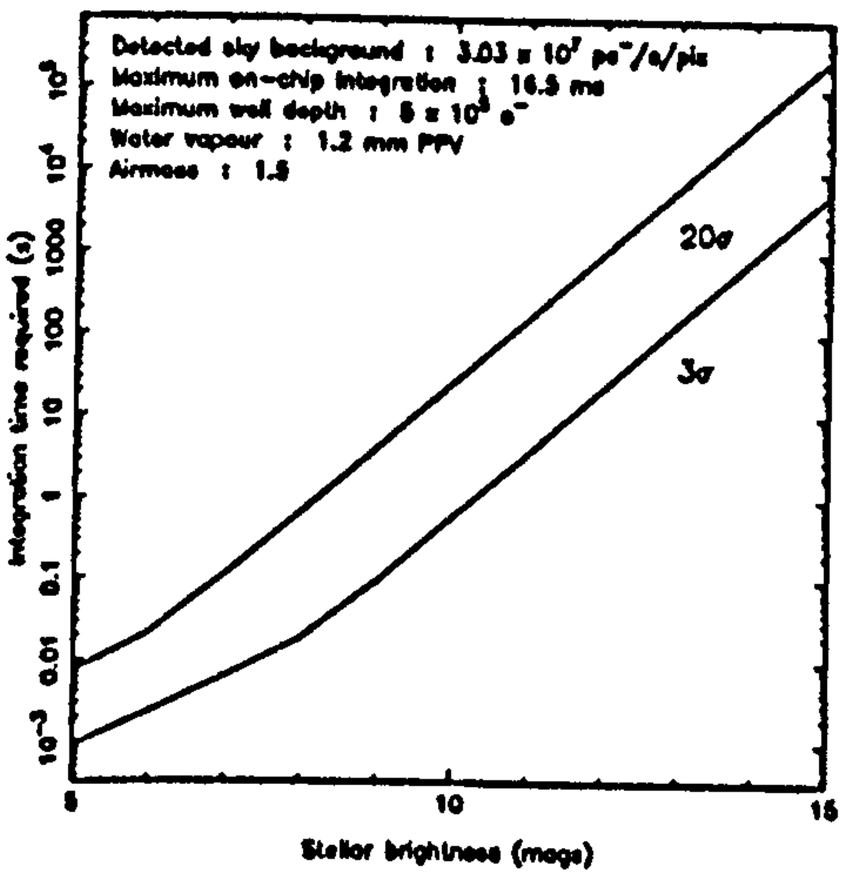


Figure 4.48: L′ : point : 0.6″/pix

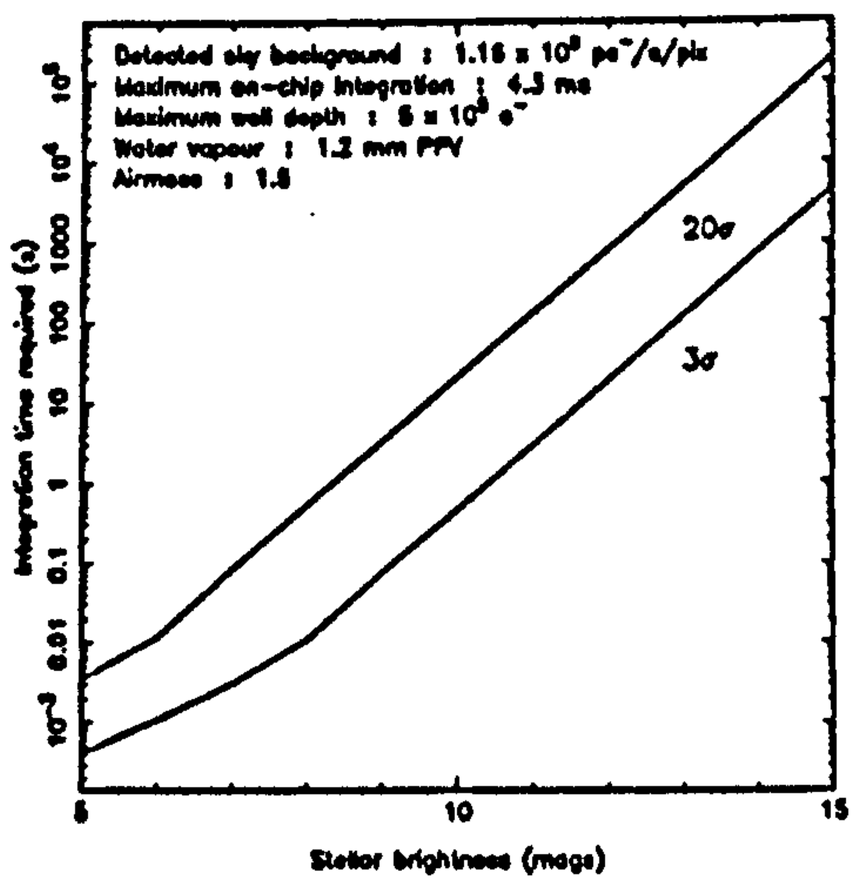


Figure 4.49: L′ : point : 1.2″/pix

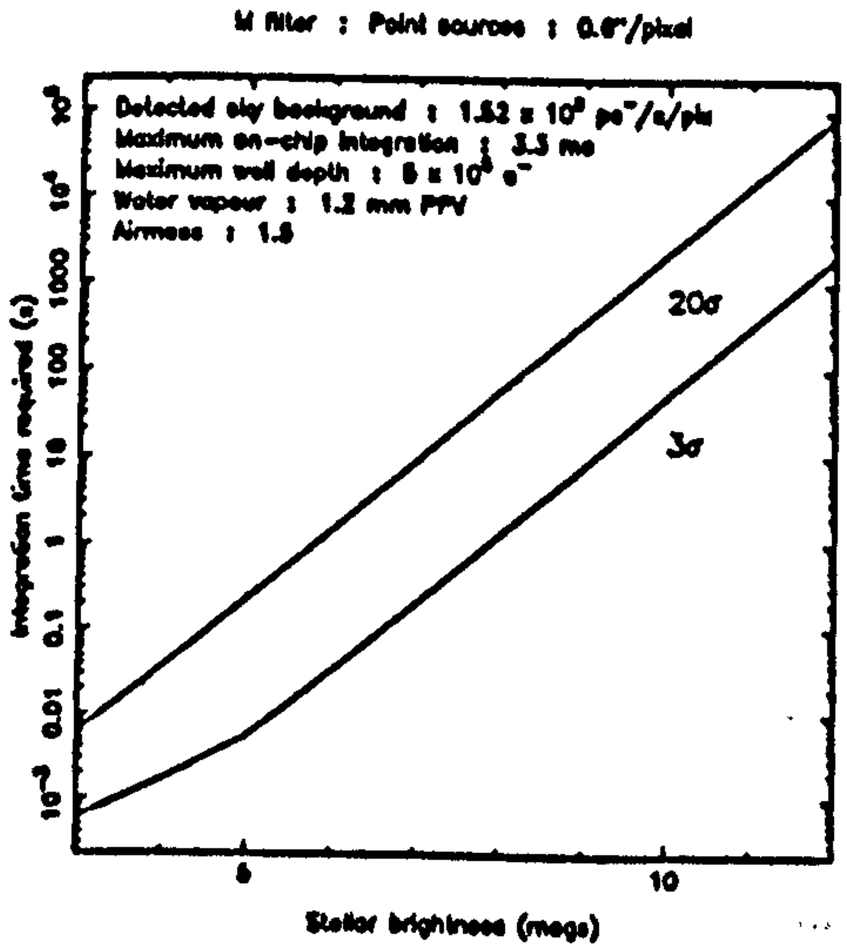


Figure 4.50: M : point : 0.6″/pix

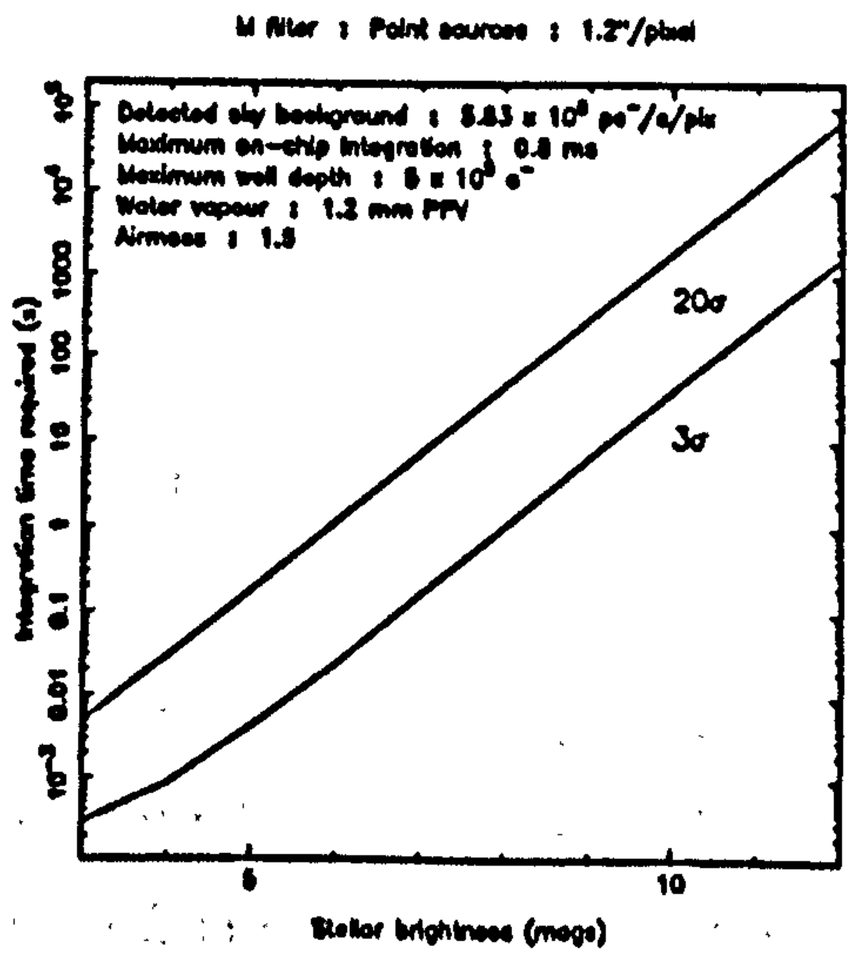


Figure 4.51: M : point : 1.2″/pix

peak pixel brightness in the 1.2″/pixel mode is roughly twice that for the 0.6″/pixel mode. As the background goes up a factor of 4 between the modes, the signal to noise will remain fairly constant on average.

4.11 Simulated images

In order to establish some feeling for the possible quality of infrared camera data, we present images that realistically simulate the appearance of star fields and extended emission regions for a range of likely parameters.

4.11.1 Point source simulations

As discussed near the beginning of this chapter, we use simple catalogues to hold information on simulated star fields. We have created one small catalogue for the purpose of demonstration, which we shall use for our point source simulations. The catalogue contains 15 stars, 10 of which lie within a central region of $\sim 35'' \times 35''$. The stars are shown in figure 4.52. The left hand image is a noise free representation of a 62×58 array of 0.6″ pixels, with a seeing FWHM of 1.5″. This image shows the stars scaled linearly at J, and the right hand image is a key to the J magnitudes. All but one of the stars have been given the same magnitude through each filter, in order to simulate normal foreground stars. We have used a range of $15^{m}.0$–$20^{m}.0$, with one brighter star at $10^{m}.0$ in the top right hand corner. The source at the centre of the image has a faint J magnitude of $20^{m}.0$ — however, we have simulated it as a highly reddened source, with colours $J - H = H - K = K - nbL = 2^{m}.5$.

For our imaging simulations, we have used similar parameters to those used for sensitivity predictions in the previous section. We have used a dark current of 100 e^{-}/pix/sec, a read noise of 375 e^{-} RMS, background photocurrents as predicted for IRCAM on the UKIRT at an airmass of 1.5, a water vapour of 1.2 mm, and a telescope temperature of 275 K. We have predicted the point source photocurrents through each filter from the empirical Vega model. We have assumed perfect flat fielding, i.e. we assume the residual noise after flat fielding is due to shot and read noise terms alone, as demonstrated to be true for the SBRC over a range of situations in the previous chapter.

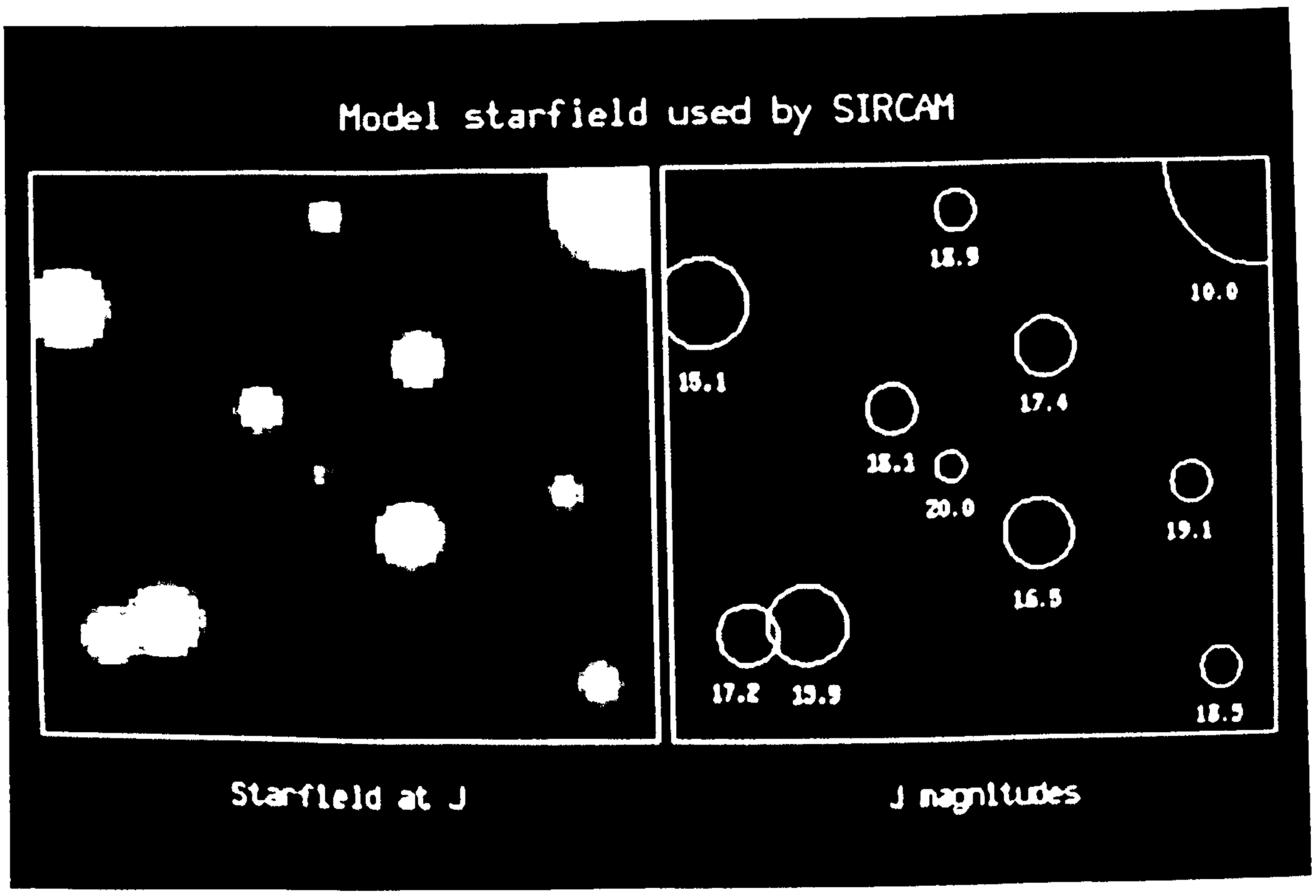


Figure 4.52: Simulated star field

We have generated the simulated images as follows. For each star in the catalogue, we have located its position in the image according to its position in the catalogue, and the specified pixel scale and image centre relative to the catalogue coordinates. From its tabulated brightness, we calculate the total number of photoelectrons per second detected at the array. We then distribute this number according to the Lorentzian seeing profile and virtual pixel technique as described earlier. At this point, we have an image containing the number of photoelectrons detected per second for each pixel in the image.

From the calculated background photocurrent per pixel, we calculate the time taken to reach half well ($5 \times 10^5\,e^-$), at which point the system is background limited, and we enforce a read out. From this time, and the total integration time to be used, we calculate the number of source and sky exposures to be made. As previously, we assume that the source and sky frames are taken in 'real time', whilst the corresponding dark current frames are taken at dusk or dawn. We assume equal numbers of co-adds for all three types of frame. In the case where the background photocurrent is low, and does not reach half well in 300 seconds, we also enforce a read out.

From the source and background photocurrents, the number of co-adds, and the

detector parameters, we then scale the source flux according to total integration time, and calculate the residual noise left after flat fielding for each pixel in the array, according to equation 3.96. Then for each pixel, we generate a 'noisy' value, calculated from the source signal and the RMS deviation, using a normal distribution.

We have displayed the resulting images close to the noise, at -1σ to $+5\sigma$ about the mean sky level. Although this reduces the visible dynamic range and enhances the noise, we have done so in order to emphasise to lowest visible levels of source flux in each case.

We have simulated images of our model star field at J, K, and nbL, in order to visualise the differences between low, intermediate, and high background imaging in the 1–5 μm region. For each filter we have simulated the results obtained after 1 minute and 1 hour total integration time. The shorter integration times at J and K are effectively read noise limited, as the background does not fill to half well in the 30 seconds allocated to the source frames. All the other images shown are background limited.

Figure 4.53 shows the field at J in the 0.6″/pixel mode after 1 minute and 1 hour. After 1 minute, we are clearly detecting stars at $16^{m}.0$–$17^{m}.0$, albeit faintly. All the stars are visible after 1 hour, including the source at J = $20^{m}.0$. Whether this latter detection would provide anything but very crude photometry is doubtful.

At K, the background has risen by a factor of 6 over that seen at J, whilst a star of equal magnitude through both filters will result in $\sim$ 33% fewer detected photoelectrons per second at K. These factors combine to lower the sensitivity relative to J, for normal, unreddened sources. Indeed, after 1 minute, figure 4.54 shows us that we are only marginally detecting stars at $16^{m}.5$, and after an hour, stars at K = $19^{m}.0$ are barely visible. Note however that our reddened central source, faint at J, has risen to K = $15^{m}.0$, and easily detected after only 1 minute.

Figure 4.55 shows the simulated nbL images. The background at nbL is $\sim$ 130 times brighter than at K, whilst the flux from stars with similar magnitudes at K and nbL have a detected flux ratio of $\sim$ 18 : 1, mainly due to the narrow spectral bandwidth of the nbL interference filter ($\sim$ 1.8%). Only the bright star at $10^{m}.0$ is visible after 1 minute. Our reddened source has risen to nbL = $12^{m}.5$, and is visible after 1 hour.

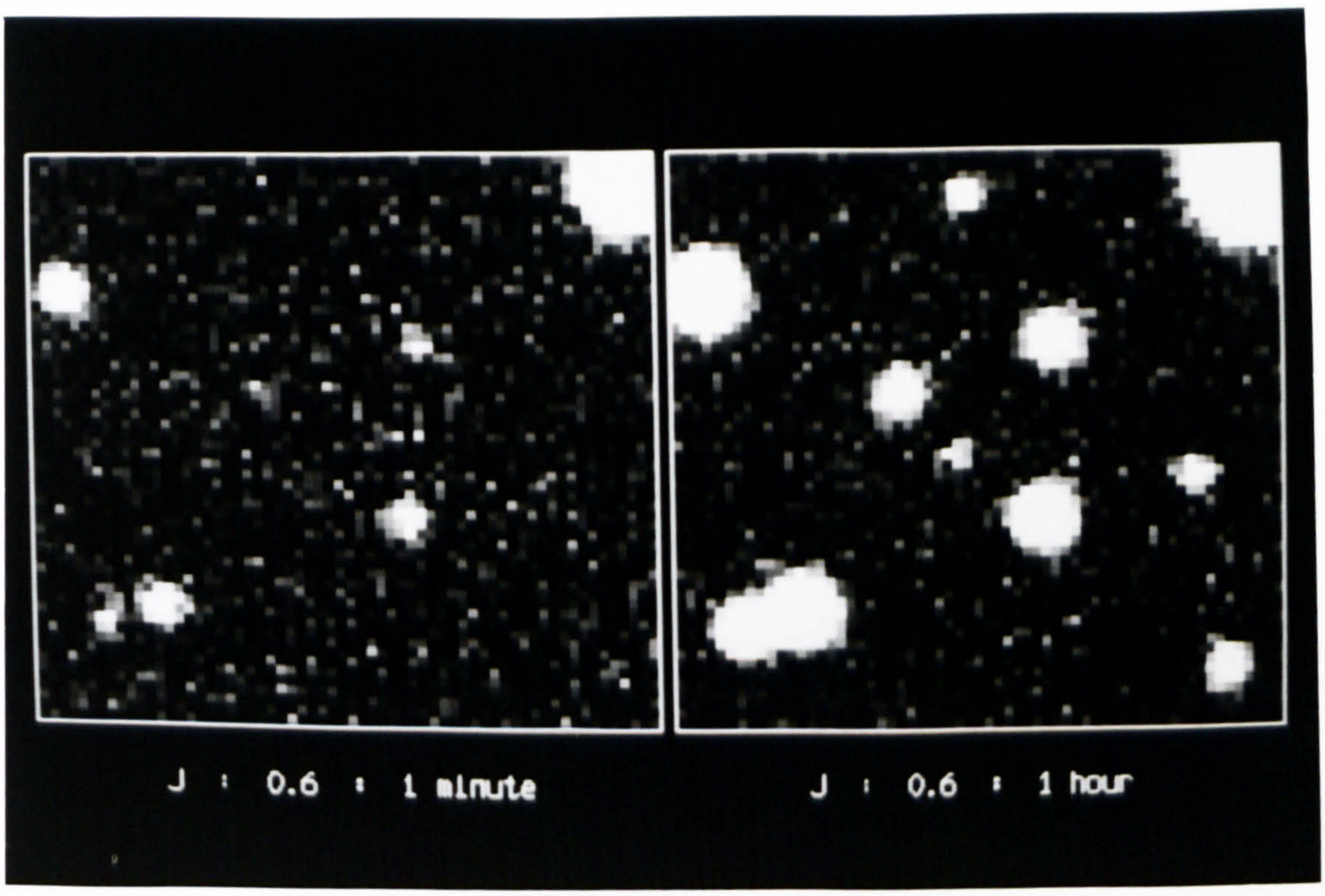


Figure 4.53: Simulated star field at J : 0.6″/pixel

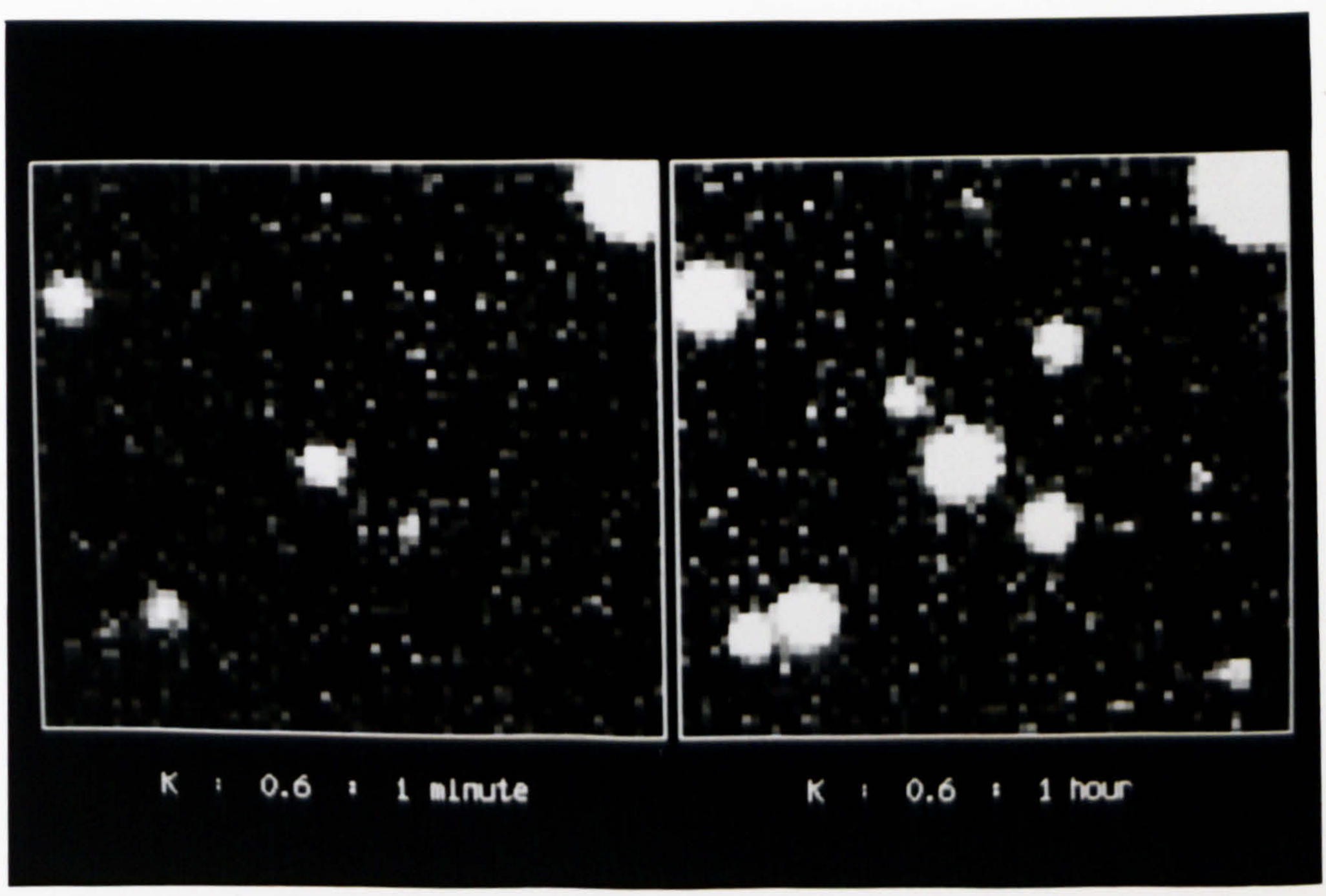


Figure 4.54: Simulated star field at K : 0.6″/pixel

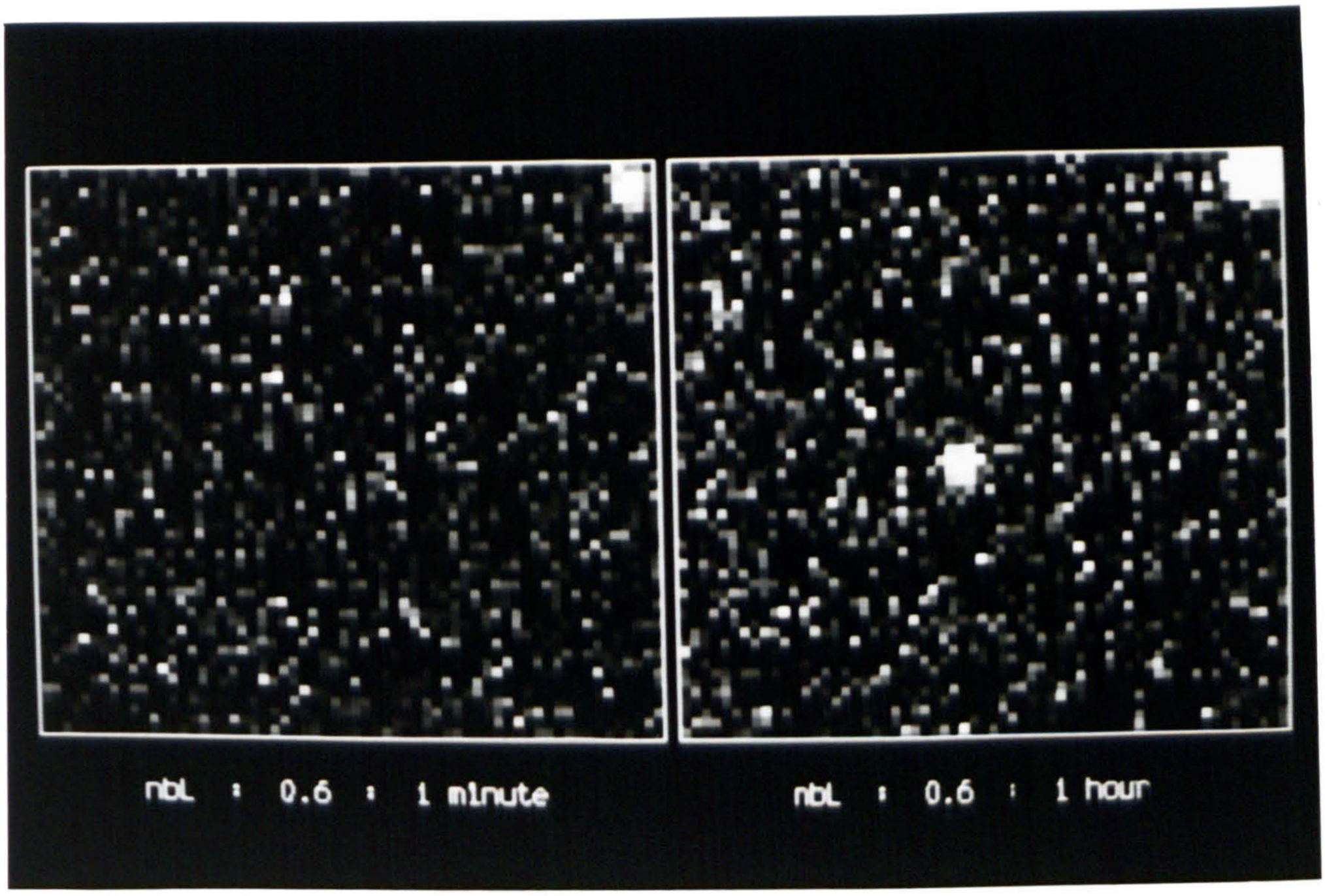


Figure 4.55: Simulated star field at nbL : 0.6″/pixel

Finally, as a rough guide to the quality and sensitivity of the lower spatial resolution imaging modes of IRCAM, figure 4.56 shows the result of 1 hour integrations at K on the same cluster, but now in the 1.2″/pixel and 2.4″/pixel modes. As suggested previously, there are no dramatic changes in sensitivity to point sources visible in the 3 modes, as may be seen by cross checking some of the fainter stars (K $\sim 19^{m}.0$). Any gains in total spatial coverage appear to be offset by degradation in image quality, even in the 1.2″/pixel mode, and in the 2.4″/pixel mode, the fainter known stars are mimicked by single pixel statistical deviations in the background, making their certain detection difficult.

Thus, by cross-referencing between the simulations shown in these images and the point source sensitivity predictions given in the previous section, a reasonably realistic assessment may be made of how faint a point source can be reliably detected in a given integration time, for the 0.6″/pixel mode at least.

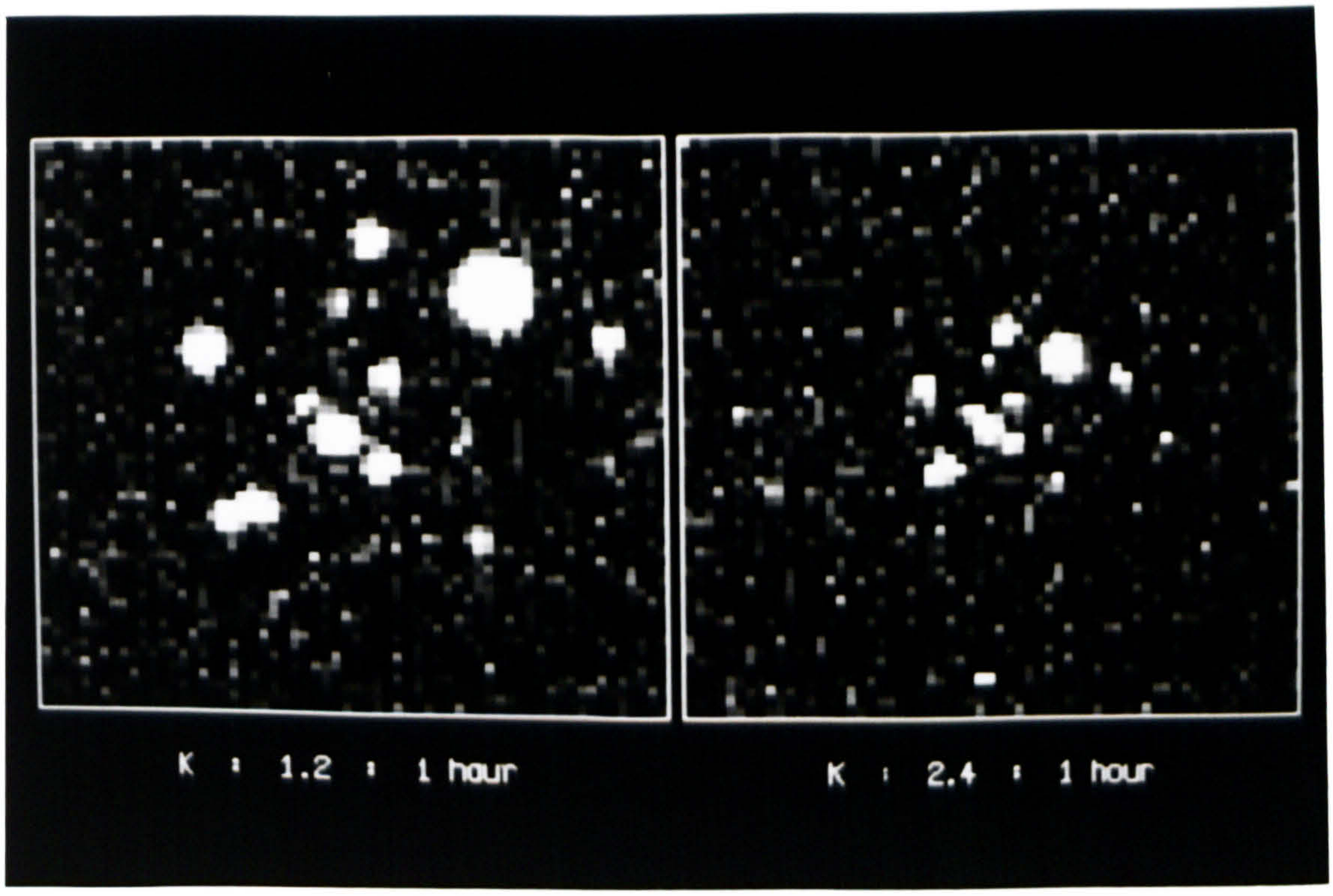


Figure 4.56: Simulated star field at K : 1.2″ and 2.4″/pixel

4.11.2 Extended emission simulations

In a similar way to that detailed above, we have also simulated some extended source images, replacing point sources with a model extended emission structure as shown in figure 4.57. A small section of a much larger simulated extended emission 'region' is seen here at 0.6″/pixel, and with 1.5″ FWHM seeing. The image is displayed logarithmically, and the interval between the numbered contours is a factor of $\sqrt{10}$ in surface brightness. Thus the dynamic range between contours 1 and 5 is a factor of 100.

We have scaled this image to simulate Brγ emission at 2.166 μm. We have assigned contour 1 to a surface brightness of $3 \times 10^{-23}\,\mathrm{W/cm^2/\,''^2}$, contour 2 to $10^{-22}\,\mathrm{W/cm^2/\,''^2}$, and so on. From the SIRCAM software, we have calculated that a surface brightness of $3 \times 10^{-23}\,\mathrm{W/cm^2/\,''^2}$ corresponds to a detected photocurrent of $2.2\,\mathrm{e^-/sec/pixel}$ in the 0.6″/pixel mode through the narrow band Brγ filter at 2.166 μm.

We use a predicted sky background of $455\,\mathrm{e^-/sec/pixel}$ for this mode, close to the actual background flux seen by IRCAM when imaging at this wavelength through a Fabry-Pérot étalon (McLean, private communication). All other parameters used are as for the point source simulations above, and the resulting images were simulated with

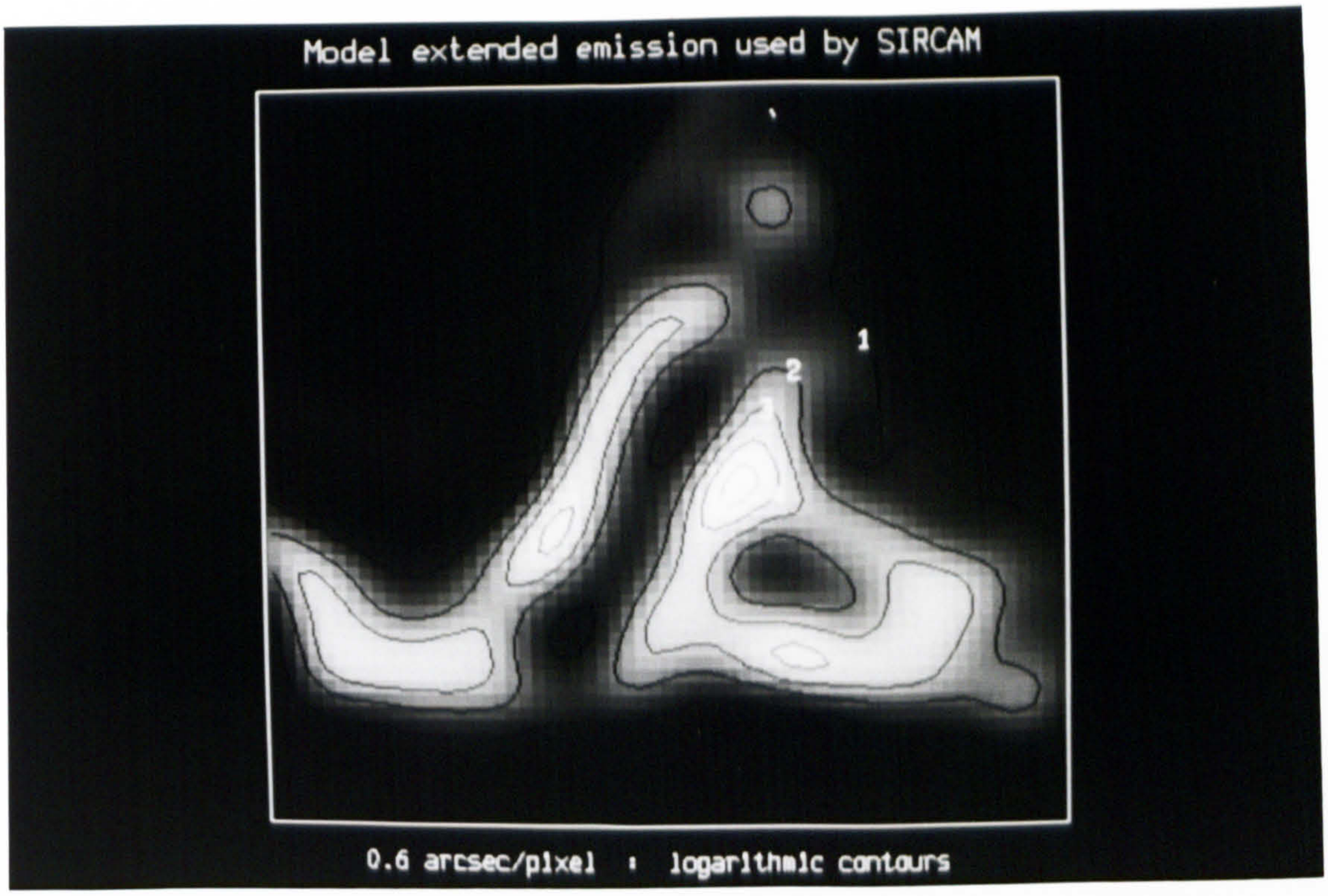


Figure 4.57: Model extended emission region

the same techniques.

Figure 4.58 shows the region after 1 minute and 1 hour, and both images are displayed linearly near the noise. After 1 minute, flux from contour 3 (3×10^{-22} W/cm^2/$\square''$) is just visible, and contours 4 and 5 (1 and 3×10^{-21} W/cm^2/$\square''$ respectively) are clearly seen. After 1 hour, contour 2 (10^{-22} W/cm^2/$\square''$) is easily seen in the small knot to the north, whilst flux from contour 1 (3×10^{-23} W/cm^2/$\square''$) is vaguely visible.

Figure 4.59 shows the effect of increased pixel size on extended emission sensitivity, with the same region shown again at Brγ after 1 hours integration, in both the 1.2$''$ and 2.4$''$/pixel modes. Clearly, the sensitivity is enhanced at the larger pixel scales, and the lower contour levels are better detected. Numerical examination of the images reveals that the signal to noise in the faintest emission improves as the square root of the pixel area, as expected. In the 2.4$''$/pixel simulation we have crudely indicated the outer edge of the unvignetted field of $\sim 120''$ seen by IRCAM in this mode. The trade offs between spatial coverage, low surface brightness sensitivity, and spatial resolution are clearly seen in these images.

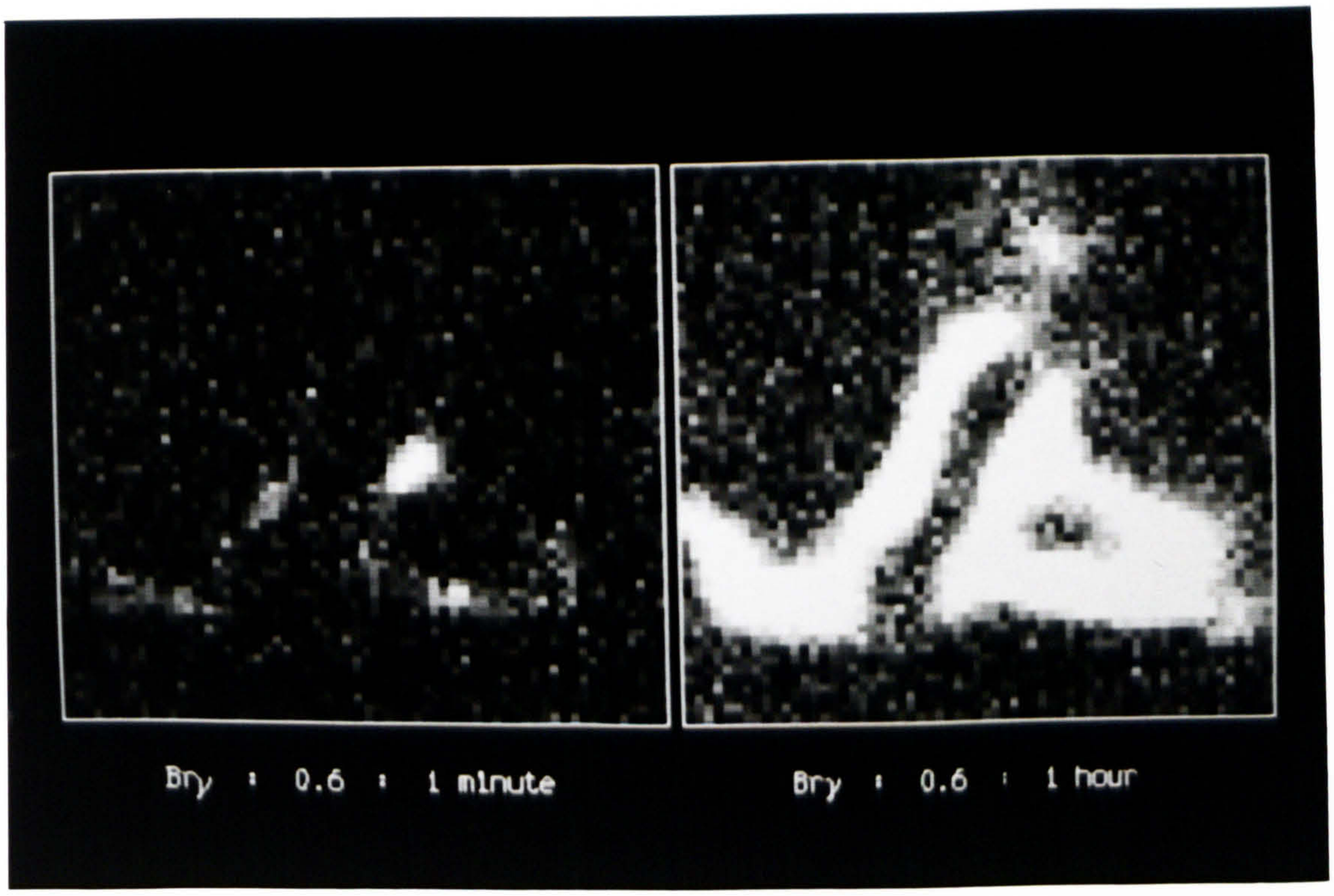


Figure 4.58: Extended source emission at Brγ : 0.6$\widehat{''}$/pixel

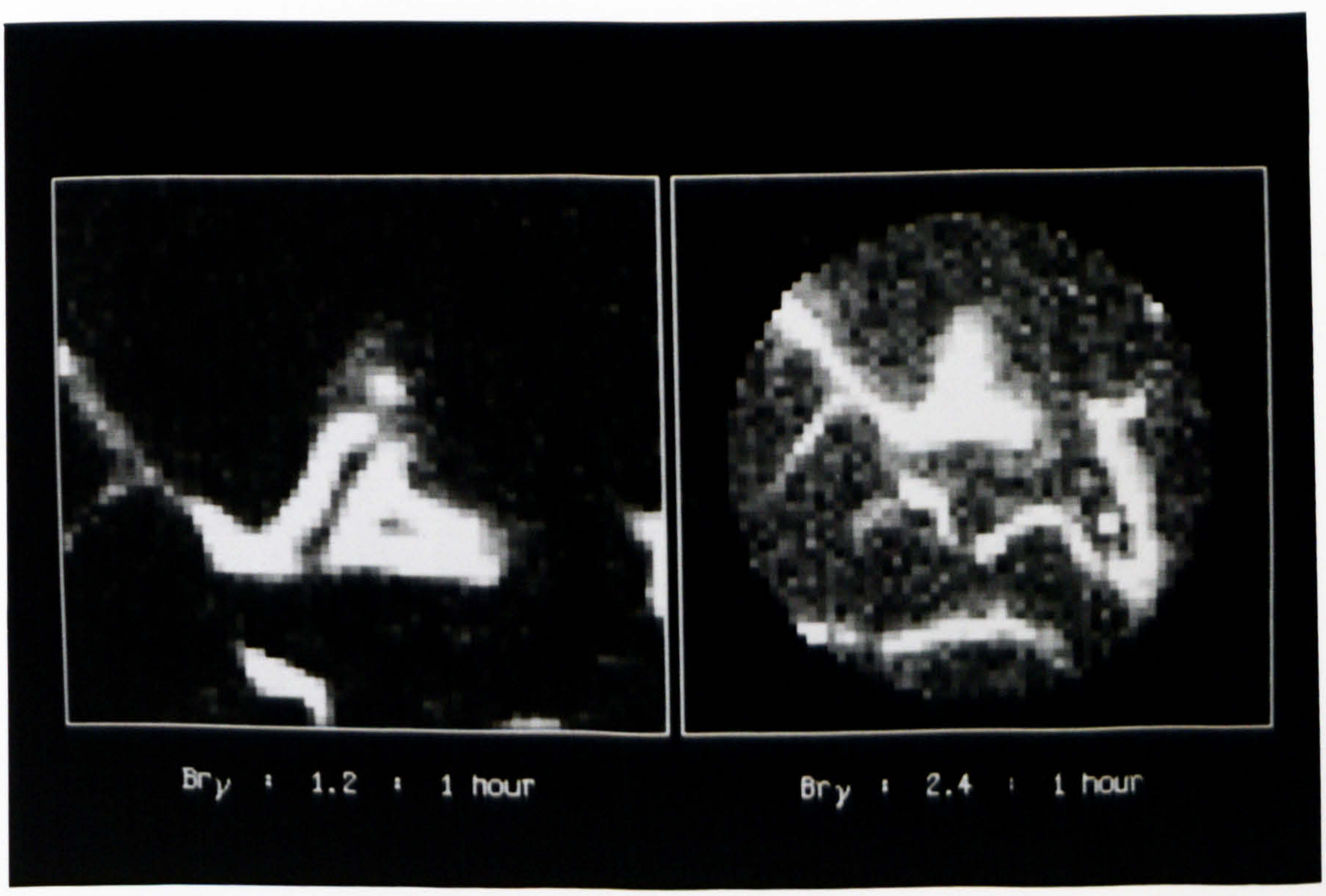


Figure 4.59: Extended source emission at Brγ : 1.2$\widehat{''}$ and 2.4$\widehat{''}$/pixel

4.12 Undersampling

In this section, we shall discuss some of the problems that may arise when using the 1.2″ and 2.4″/pixel modes of IRCAM, i.e. those modes that do not fully sample the seeing. We have previously seen how much the brightest pixel from a star of given brightness may vary in a heavily undersampled mode, depending upon where the star centre lands relative to a pixel centre. We shall now examine the effects of fill factor on photometric accuracy in the same situation, and the effects of aliasing have on information content when an array heavily undersamples the scene. We discuss the technique of micro-scanning as a solution to these problems.

4.12.1 Photometric errors

In the heavily undersampled modes, most of the flux from a point source can fall within the boundaries of a singe pixel. As we have seen previously, this leads to a large range in the measured peak pixel signal for stars of equal intensity, depending on exactly where the star centre falls relative to the centre of a pixel.

In principle however, we should be able to add together the signals from several adjacent pixels and retrieve the total brightness, regardless of this positioning. In practice though, the area fill factor must be accounted for. Incoming photons will go undetected if they fall on to the dead space which surrounds each pixel. In undersampled modes, the effects of the dead space on the total flux lost will differ according to the star centre relative to a pixel centre, leading to an inherent photometric inaccuracy. For well sampled modes, the fraction of the total flux lost will be relatively constant, regardless of where the star centre lands.

We have simulated the effects of dead space on photometry. To do so, we have employed the virtual pixel technique, as discussed earlier. We subdivide each physical pixel into an array of virtual pixels, calculate the signal for each virtual pixel, and then sum to find the total signal for the physical pixel. To simulate dead space, we set the relevant fraction of virtual pixels around the edge of the physical pixel to have zero quantum efficiency.

Table 4.5 shows the results of simulations for each of the IRCAM imaging modes

Pixel scale	Fill factor	Centre		Cross	
		Peak	Total	Peak	Total
0.6″/pix	100%	0.2616	1.0000	0.1665	1.0000
	90%	0.2387	0.9025	0.1509	0.9032
	50%	0.1442	0.5103	0.0884	0.5131
1.2″/pix	100%	0.6663	1.0000	0.2402	1.0000
	90%	0.6317	0.9217	0.2124	0.8848
	50%	0.4511	0.5841	0.1081	0.4476
2.4″/pix	100%	0.9610	1.0000	0.2507	1.0000
	90%	0.9503	0.9819	0.1982	0.7929
	50%	0.8597	0.8690	0.0516	0.2065

Table 4.5: Effect of fill factor on photometric accuracy

(0.6,1.2,2.4″/pixel). We have assumed a Lorentzian seeing profile with a FWHM of 1″. The extreme cases are where the star centre falls at the centre of a physical pixel (centre), and where the star centre falls at the intersection of four pixels (cross). We calculate the fraction of the total input flux that is seen in the brightest pixel, and then the fraction of the total input flux that can be retrieved by binning several adjacent pixels. The differences in the latter number for the two extreme cases will be a measure of the maximum photometric error that may be expected. The results shown are for three area fill factors, namely 100%, 90%, and 50%.

We can see that the 0.6″/pixel mode gives excellent results, even with a low area fill factor of 50% — the maximum photometric error is as low as 0.3%. The 1.2″/pixel case is not so good. With a fill factor of 90%, a maximum photometric error of 4% can be expected, climbing to 14% error for a fill factor of only 50%. However, as expected, it is the 2.4″/pixel mode results which are the worst. With a fill factor of 90%, the maximum photometric error predicted is almost 20%, and this rises to almost 70% when the fill factor is 50%.

The SBRC array has an area fill factor of about 90% [116], and thus we can see that whilst the 0.6″ and 1.2″/pixel modes give reasonably accurate photometric results, the 2.4″/pixel mode has an inherent photometric error of up to 20%.

In chapter 2, we briefly mentioned the technique of micro-scanning as a method

of compensating for low area fill factor [84]. This essentially reduces the photometric error by making many exposures of the source with sub-pixel offsets between each, and averaging the photometric results. We shall discuss this technique further below, with regards reducing the effects of aliasing.

4.12.2 Aliasing

Another problem that arises when using array detectors, particularly at low spatial resolutions, is aliasing. This effect is due to poor sampling, and can lead to spurious effects appearing in an image. If a given system samples a scene at a spatial frequency f (i.e. the pixel-to-pixel pitch corresponds to that spatial frequency), then no information with spatial frequencies greater than $f/2$ can be unambiguously obtained from the scene, due to the requirements of the Nyquist sampling theorem. Another way of saying this is that if information is wanted at a given spatial frequency, then the scene must be sampled twice as frequently. If the sampling frequency is p_s, then the maximum frequency at which information can be obtained is $N_f = 1/2p_s$.

If there is spatial information in a scene at a higher frequency than N_f, then aliasing will inject that information back into the frequency range 0–N_f as spurious detail. The worst aliasing effects occur when there is strong regular spatial information at frequencies $\geq N_f$ — this is rarely encountered in astronomy due to the irregular structure of the observed sources, but secondary effects of aliasing, such as poor positional accuracy, can still be noticeable [91].

One way of preventing aliasing is to band-limit the incoming frequency spectrum, by blurring the image. This may be achieved by slight defocusing of a lens, but is conveniently achieved in astronomy by atmospheric turbulence, the seeing. Thus, if the seeing FWHM is $1''$, then with a pixel scale of $0.5''$/pixel, the maximum information content is obtained, and aliasing is avoided. Of course, deconvolution techniques exist to improve the spatial resolution further (e.g. CLEAN, MEM, speckle), but these are discussed elsewhere.

However, lower spatial resolutions may be required, in order to maximise sensitivity to low surface brightness sources. In these cases, the pixel scale is generally much larger than the seeing FWHM, and the scene is undersampled. Care must be taken to avoid

aliasing, and also that no conclusions are drawn about features smaller than twice the size of a pixel. Thus, in the 2.4″/pixel mode, no information is available at scales smaller than 4.8″, and the image should be smoothed accordingly to prevent spurious identifications.

As mentioned briefly above, the technique of micro-scanning may be used to compensate for a low area fill factor. The same technique may also be used to overcome aliasing, and to improve the spatial resolution.

Using a small mirror scanned using a piezo-electric stack, Bradley and Dennis [24], and Dann *et al.* [31] have demonstrated the significant reductions in aliasing and improvements in spatial resolution that can be achieved with simple 2 × 2 and 3 × 3 micro-scans. Obeying the sampling theorem, and sampling ≥ 2 times per pixel pitch, the aliasing is considerably reduced. The Nyquist frequency (i.e. a measure of the observable detail) is also increased by micro-scanning and suitable post-detection image processing. A 4 × 4 micro-scan procedure could in principle allow seeing limited details to be retrieved from a sequence of 2.4″/pixel images.

In the background limit, a 4 × 4 micro-scan of a region in the 2.4″/pixel mode should achieve the same spatial resolution, sensitivity, and total spatial coverage in the same time as would be needed in the 0.6″/pixel mode. This would be achieved by off-line processing of the micro-scanned images onto a higher spatial resolution grid. Alternatively, the images could be co-added in such a way that the spatial resolution was maintained at 2.4″/pixel, but retaining the factor of four gain in signal to noise for extended sources over the 0.6″/pixel mode.

However, at short wavelengths and high spectral resolution, the 0.6″/pixel mode may well be system (read, dark) noise limited, whilst the 2.4″/pixel mode, with its larger aperture per pixel, may well be photon background limited. In this case, a given integration time will bring the same spatial resolution and area coverage in both modes, but a higher sensitivity to the micro-scanned 2.4″/pixel mode. This gain may be considered as a spatial analogy to the multiplex gain achieved by Fourier transform spectrometer systems, which win over grating spectrometers in a system noise limited case.

Although no higher spatial resolution information would be available without us-

ing deconvolution techniques as mentioned earlier, the use of micro-scanning in the 0.6″/pixel mode would still bring benefits by reducing aliasing and photometric problems due to imperfect fill factor and intra-pixel variations, and could be useful during periods of exceptional seeing (< 1″FWHM).

The implementation of a suitable micro-scan system should be relatively easy, given the accuracy with which the chopping secondary may be controlled, and at the expense of some additional off-line processing, would bring significant improvements in photometric accuracy, imaging detail, and a possible multiplex gain in system noise limited situations, particularly to the lower spatial resolution modes.

4.13 Conclusions

In this chapter we have examined the SIRCAM system, and its application to simulations of IRCAM in the near infrared. We have seen that the background photocurrents in the 1–5.5 μm region have a very large dynamic range, and that the region may be subdivided into the 1-2.5 μm non-thermal regime, where the dominant background source is the highly structured and highly variable OH* emission from the mesopause, and the 2.5–5.5 μm regime, where thermal emission from the telescope and warm optics dominates, with thermal emission from the atmosphere becoming increasingly important at longer wavelengths. A high spectral resolution model of the OH* emission has been derived from published literature, and has been shown to match the observed night sky spectrum very closely, across the limited region available for comparison. The nature and scale of total flux and spectral structure variations have been discussed, and although the latter variations will not be too important to imaging, they will deserve considerable study as infrared arrays are incorporated into high resolution spectrometers.

We have predicted sensitivity figures for IRCAM across a wide range of its possible observational configurations, and have shown that very faint point and extended emission sources should be detectable in reasonable integration times, mainly through the use of long enough on-chip integrations to become background limited. We have demonstrated by means of these sensitivity figures and realistically simulated images the trade offs that arise between spatial resolution, sensitivity, and total spatial coverage that arises when the pixel scale is varied, from seeing limited resolution at 0.6″/pixel,

to heavily undersampled imaging at 2.4″/pixel. We have also demonstrated that photometric accuracy and aliasing may pose problems in the heavily undersampled modes.

In order to confirm or contradict these predictions, we must await the results of rigorous photometric testing of IRCAM. Limited results available thus far indicate that the predicted background fluxes are within factors of 2–3 of the measured fluxes, and that small thermal leaks or slightly different filter profiles may account for the discrepancies. The predicted sensitivities to point sources and extended emission have been derived from a noise model that accurately predicts the background noise seen in IRCAM and the SBRC array, and therefore should not be too much in error. However, the limited test data available do not extend to very long on-chip integrations, nor very long total integration times, and therefore care must be taken when extrapolating sensitivity limits. However, in most cases in the non-thermal regime, IRCAM is background limited within a minute, and considerably quicker in the thermal regime. Individual background limited frames should combine to give square root improvement with time, and special observing procedures can used to ensure this gain, such as slight offsetting between frames.

Clearly however, the main advantage of infrared arrays is their large number of pixels. This allows great flexibility, and individual observers can choose to optimise their experiments by selecting, say, a high resolution imaging mode for detailed spatial examination of complex regions, an intermediate resolution mode for large area surveys with high sensitivity and limited spatial information, or a low resolution mode in order to enhance the sensitivity to low surface brightness sources.

Chapter 5

Infrared imaging of the Orion Nebula

In this chapter, we shall present new near infrared images of the Orion Nebula (M42) obtained with IRCAM on the UKIRT during the winter of 1986–1987. We shall be discussing these new images with five main purposes in mind :

- How the raw data were obtained at the telescope
- Which procedures were used to reduce the raw data to the finished images
- How the finished images were analysed to give useful numerical data
- How the new images compare with previous infrared mapping and imaging of the same region
- The novel scientific results arising from the new images

Firstly though, it must be noted that the majority of the data presented here were obtained during the early part of the instrument commissioning period, within two months of first light on the telescope. Combined with the fact that only engineering grade detectors were used, it is therefore stressed that many of the specific parameters, such as sky background and number of bad pixels for example, are no longer applicable to IRCAM. The camera has has developed and improved considerably during the following year, in addition to the installation of a series of new scientific grade detectors, and is

now more sensitive and stable than during the time when these data were obtained. Also, during the commissioning period, lessons were still being learned, and subsequent processing of the data revealed several important oversights and errors made when the data were first obtained.

However, most of the comments relating to data reduction and analysis are still relevant, and may be taken to represent the most up-to-date information. Finally and most importantly, despite the engineering nature of the data, they nevertheless represent the the highest resolution infrared views of the Orion Nebula yet seen, and have revealed several new conclusions of astronomical importance.

We begin by briefly discussing our motives for obtaining these data. Then we continue with a detailed examination of the observing procedures, and data analysis and reduction techniques used to obtain a seeing limited $5.3\hat{}' \times 4.7\hat{}'$ image of the Orion Nebula at $2.2\,\mu$m, through the K broad band filter, with particular emphasis on the techniques used to make a large, smooth mosaic of 126 individual frames. Then we briefly discuss an image of the same region made at $1.6\,\mu$m (H), which covers a smaller area and is of a lower quality than the K image due to instrumental problems, and also an image of a limited section of the region made through a narrow band interference filter at $3.6\,\mu$m.

In the subsequent scientific analysis of these data, we shall concentrate mainly on the K image. We compare our new images with previous mapping and imaging of the region in the near infrared, particularly in the region around the bright infrared source BN to the north-west of the well-known Trapezium stars at the centre of the images. Accurate astrometry is derived from the K image and specially commissioned optical far red Schmidt plates. Photometry of the dense cluster of stars is discussed, and a K frequency function is presented and compared with one previously derived for a more limited section of the cluster. We discuss several embedded infrared point sources outside the region around BN, some of which were previously known, and some newly discovered. We particularly concentrate on several reddened point sources clustered near the core of a southern component of the molecular cloud OMC-1, and demonstrate which of these sources are most likely responsible for the luminosity of the molecular clump. We examine embedded point sources found near several of the known Herbig-Haro objects in the region, and suggest that these point sources may be their source of

energy. We examine the spatial coincidences between members of a recently discovered family of compact radio sources and infrared point sources, and discuss the apparent physical extent of the infrared objects with seeing limited resolution. Finally, we discuss future work needed in this region as a consequence of the results presented here.

5.1 Motives

As mentioned at the beginning of this chapter, the initial motivations for our study of the Orion Nebula were of an instrumental nature. As the field of view of the current generation of infrared detectors is limited compared to that of CCDs, many infrared imaging studies will require multiple image mosaics, and as part of the commissioning process we wished to examine the techniques required to obtain a large number of images of a given source, and then to combine the reduced images into a single large calibrated mosaic.

This involves problems of an observational nature, including the frequency of sky referencing and sampling strategies, and problems of an analysis nature, including intensity matching and suitable mosaicing algorithms.

The Orion Nebula was a natural choice for such an experiment, being amongst the most heavily studied regions in the sky. The HII region M42 and its central cluster of bright OB stars known as the Trapezium signpost the Orion Nebula to the optical astronomer, and much work has been done, particularly in emission lines associated with the HII region. The discovery in the infrared of the heavily reddened point source BN [19] to the north-west of the Trapezium led to the subsequent discovery of the much younger cluster of sources centred on IRc2, deeply embedded in the molecular cloud OMC-1, behind the HII region, and only visible at wavelengths longer than 1 μm. This region is often referred to as BN-KL, even though the Kleinmann-Low nebula has since been resolved into a cluster of point sources, extended sources, and reflection nebulosities [166]. We shall use the term BN-KL throughout this chapter to denote the square arcminute surrounding the presumed power source for the region, IRc2, just a few arcseconds to the south-east of BN. The foreground HII region, the background molecular cloud, the optical Trapezium cluster, and the embedded infrared cluster of newly formed stars have been intensively studied at all wavelengths, from x-ray to radio,

and most particularly in the optical and infrared. An excellent summary of these data, and how the Orion Nebula fits into the larger scale picture of star formation in Orion may be found in Goudis' book "The Orion Complex : A case study of interstellar matter" [52].

As the Orion Nebula is the nearest and brightest of the galactic star forming regions, new observational techniques are frequently tested there before harder targets are attempted. As infrared instrumentation has progressed, the complex region surrounding BN and IRc2 has been subjected to more detailed scrutiny, and it is natural that the new technique of near infrared imaging with detector arrays should be used to probe this region at higher spatial resolution than previously possible with single element detectors. During the analysis of the images presented here, we shall be frequently referring to three papers, each of which represented the highest spatial resolution near infrared studies of this star forming centre in the Orion Nebula on publication.

The first of these, from 1982, is by Lonsdale *et al.* (LBLS) [90], who used a tangent beam chopping technique with a $3.5''$ aperture at H, K, and L, over a region of $40'' \times 40''$, centred on BN-KL. This study revealed a dense cluster of twenty six point sources, approximately half of which could not be identified with foreground optical stars. In the second paper, Hyland *et al.* (HABW) [69] used a DC raster scanning technique at K, to cover a strip of $50'' \times 120''$ centred on BN, again using a $3.5''$ diameter beam. However, this DC mapping technique gave them increased sensitivity to extended sources, deliberately removed by the tangent beam chopping technique of LBLS. From their work, HABW concluded that many of the infrared point sources discovered by LBLS could not be confirmed, because they were either unseen or appeared as extended nebulosity in the $2.2\,\mu$m DC raster map. HABW also covered enough stars in their maps to derive a K frequency function, which we shall discuss later.

The final work we shall refer to, due to Allen, Bailey, and Hyland (ABH) [8], also contains data obtained with the DC raster scanning technique, and is a superset of the data presented by HABW. In their 1984 Sky and Telescope article, ABH presented a multicolour image made by superimposing J and K maps of the $5' \times 5'$ region surrounding the Trapezium, with a spatial resolution of $3.5''$. They also showed higher spatial resolution multicolour images of the BN-KL region and the Trapezium at H, K, and L′, the bright bar at J, H, and L′, and an image encompassing the Trapezium

and BN-KL in two micron continuum and emission lines of ionised (Brγ) and molecular (v=1–0 S(1)) hydrogen. Despite the fact that these images represented the highest spatial resolution images of Orion prior to the advent of imaging cameras, and that these images demonstrated the enormous potential of high spatial resolution, multiwavelength infrared studies, ABH have not published any quantitative results resulting from them.

Two other papers must be noted with regards high spatial resolution imaging of the BN-KL region in the near infrared. Wynn-Williams *et al.* [166] presented a 2″ resolution map of the 15″ × 15″ centred near IRc2, along with similar maps at 3.8–30 μm, with the main conclusions that IRc2 is the major luminosity source in the region ($\sim 10^5$ $L_\odot$), and that the Kleinmann-Low nebula is probably a collection of clumps around the edge of the IRc2 outflow cavity, with IRc2 as their source of luminosity. Also, the University of Rochester group have published near infrared images of the BN-KL region on several occasions, as a demonstration of their imaging camera [38]. However, no quantitative paper has been published as a result, apart from a short two page note by Woodward *et al.* in 1986 [162].

Thus, the large amount of relevant preceding work on detailed near infrared mapping and imaging of the Orion Nebula provides us with direct comparisons for the technique of infrared imaging with array detectors, of the sensitivity, the spatial resolution, the positional accuracy, and photometric accuracy. An additional factor in favour of our imaging the Orion Nebula as part of the commissioning process was that the most important central 5′ × 5′, containing the Trapezium, BN-KL, and the bright bar, could be imaged to reasonable sensitivity limits within one night.

5.2 Observational parameters

We begin this section by discussing the observational procedures used to obtain 126 individual images taken to make up a seeing limited image (0.6″/pixel) at K (2.2 μm) of the central 5.3′ × 4.7′ section of the Orion Nebula centred on θ^1 Ori C, the brightest Trapezium star. Throughout the section we shall refer to figures which show the various stages of data reduction as applied to one single source frame, taken from a relatively clean region to the north-east of the final mosaic. Note that all the frames shown during the reduction stages are oriented with north up, and east to the right. Only the final

Array type	:	SBRC 62 x 58 InSb + DRO
Array number	:	FPA004
Pixel scale	:	0.625″/pixel
Wavelength	:	K (2.2 μm broad-band)
V_{bias}	:	−200 mV
V_{gate}	:	−2.8 V
Read-out rate	:	129.6 msecs
Detector temperature	:	49.6 K
Electronic gain	:	40 e^-/DN
On-chip integration time per co-add	:	2 seconds
Number of co-adds per frame	:	15
Integration time per frame	:	30 seconds
Effective integration time per pixel	:	60 seconds
Total number of frames	:	126

Table 5.1: Observational details for K mosaic of Orion

images are horizontally reversed to give the conventional north up and east left.

The observations were made on night of the 9th–10th December 1986, between 0140 and 0440 Hawaiian Standard Time (Universal Time 1140–1440, 10th December 1986), and were obtained with IRCAM on the 3.8 m United Kingdom Infrared Telescope, Mauna Kea, Hawaii. The instrumental parameters are listed in table 5.1.

Starting at the Trapezium, the centres of the following images were defined as follows. A spiral raster pattern was drawn up, in the form of a 9 × 9 image grid, inlaid with an 8 × 8 grid. Thus, 145 positions were defined, at half-array steps, as may be visualised from figure 5.1, which shows just the central 25 positions. In this manner, every position on the sky is imaged twice, and by a different pixel on the array each time. This has the following advantages :

- The integration time for any given point is doubled. If the image is background limited, a factor of $\sqrt{2}$ in sensitivity is obtained. In cases where flat-field noise may be important, imaging on another pixel will help reduce this noise, and thus increase the sensitivity.

- By imaging on different pixels, points imaged by a 'bad' pixel in one image may

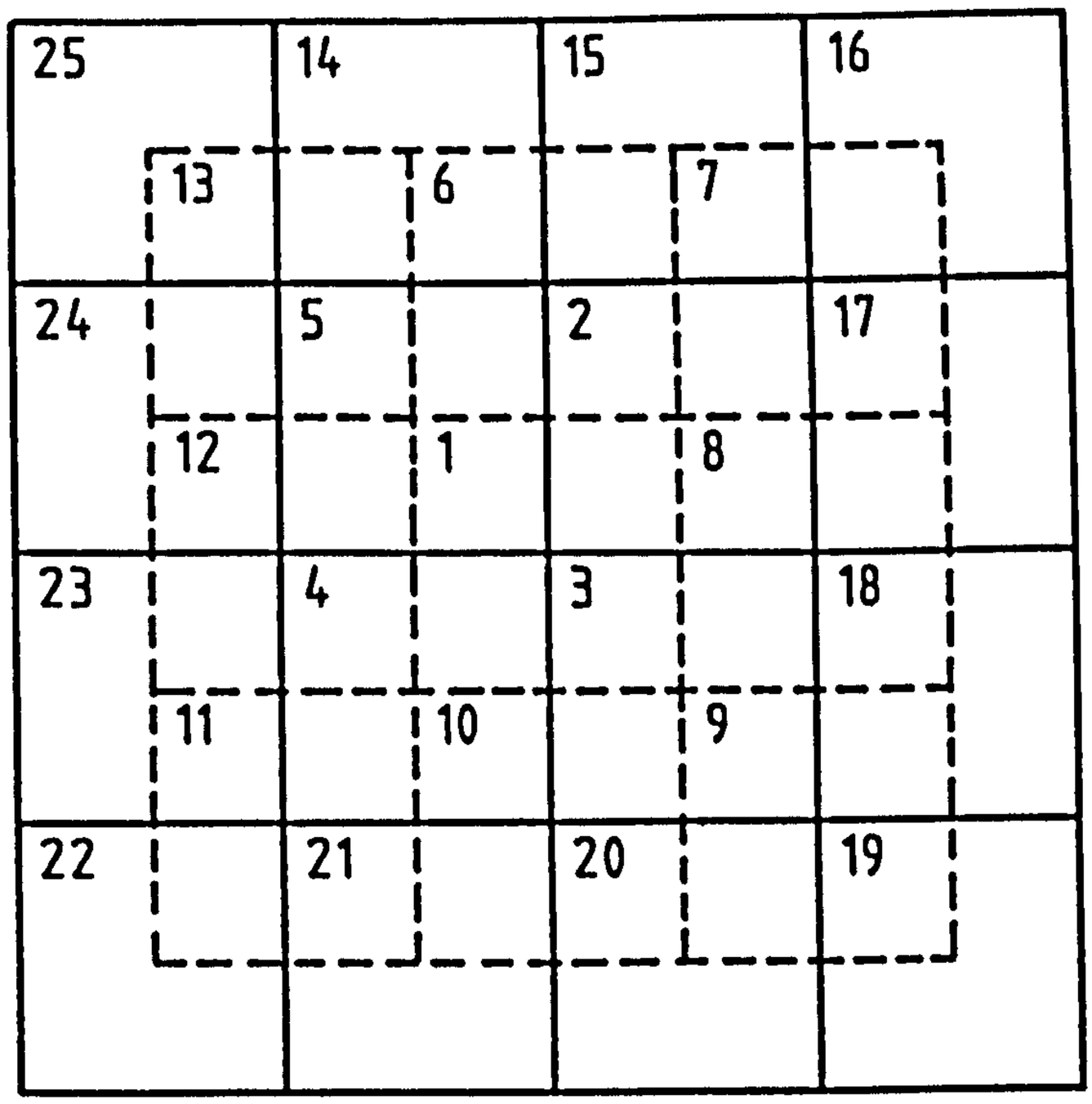


Figure 5.1: Representation of image positions in Orion K mosaic

be replaced by the data obtained from the other image, if that point is seen by a 'good' pixel in the latter image. This reduces the effect of bad pixels considerably.

- Imaging the same piece of sky twice allows for the offsets between the two images to be defined quite accurately, if (say) a point source lies in the commonly imaged region. This is important in maintaining the overall spatial accuracy.

- As the positioning error of the telescope is of the order of the pixel size, this technique goes some way to satisfying sampling criteria, as exact integer pixel offsets cannot be ensured.

From previous work, an approximate plate scale of 0.625″/pixel had been calculated. This gave a total field of view of the 62 × 58 pixel array of 38.75″ × 36.25″, and thus the steps in the raster pattern were defined to be 19″ in right ascension, and 18″ in declination. For example, the offset distance between images 1 and 2 in figure 5.1 is 19″ west and 18″ north. Due to problems with some bad rows and columns along the north and west edges of the array, these offsets were later found to be slightly larger than desirable, as discussed later.

The central (0,0) position for the mosaic was defined as the brightest Trapezium star,

θ^1 Ori C. This bright optical star was also used as a reference to check for telescope drift, using the TV guide camera. In order to flat field and sky subtract the data, a nearby piece of blank sky was required. Searching on a trial and error basis with IRCAM, a piece of blank sky was found only $-155\hat{''}, +150\hat{''}$ from the nominal 0,0 position. The area was clear of all point sources down to K = $15\overset{m}{.}1$ (see section 5.4.5.4), but was only just big enough — some stars lay just outside the camera field of view to the east and west, and occasional slight telescope drifts and positioning errors resulted in one particular star being seen at the edge of several sky frames. This is discussed further in section 5.3.3.

The sequence of observations was as follows. Firstly, a blanked-off frame was taken with the same on-chip integration time and number of co-adds as used for the following imaging observations. This measured the dark current. After an initial sky frame, six source positions were imaged before another sky frame was taken. This sequence was followed, with a sky frame taken after every six source frames — approximately eight minutes elapsed between sky frames. After every thirty source frames, another dark current frame was taken, an interval of about forty minutes. The spiral pattern was followed until position 85, when, with one hour left before the region set below the telescope limit, it was decided to obtain data at the remaining positions thought to be most interesting, namely positions 92–113 and 127–145, though not in that sequence. Thus, a total of 126 positions were observed, out of the defined set of 145, the main areas omitted being to the north and west of the Trapezium. Although the cluster and nebulosity extend out of the image to the north and west, it was felt more important to completely cover the Orion 'bright bar', to the southeast of the Trapezium.

Examples of raw dark, sky, and source frames are shown in figures 5.2 to 5.4 respectively, each scaled to show the important structure. In the dark frame, we can see that there are numerous scattered hot pixels, and a diagonal line of them, known to be associated with a crack in the detector material. On the left edge of the array, a bright region can be seen, which is due to excess dark current. This is thought to be due to localised heating of the detector array by off chip electronics. The only periodic structure seen in the dark frame is a slight odd-even effect in the columns, due to the slightly mismatched gains of the two output FETs.

The sky and source frames are very similar to each other, but notably different to

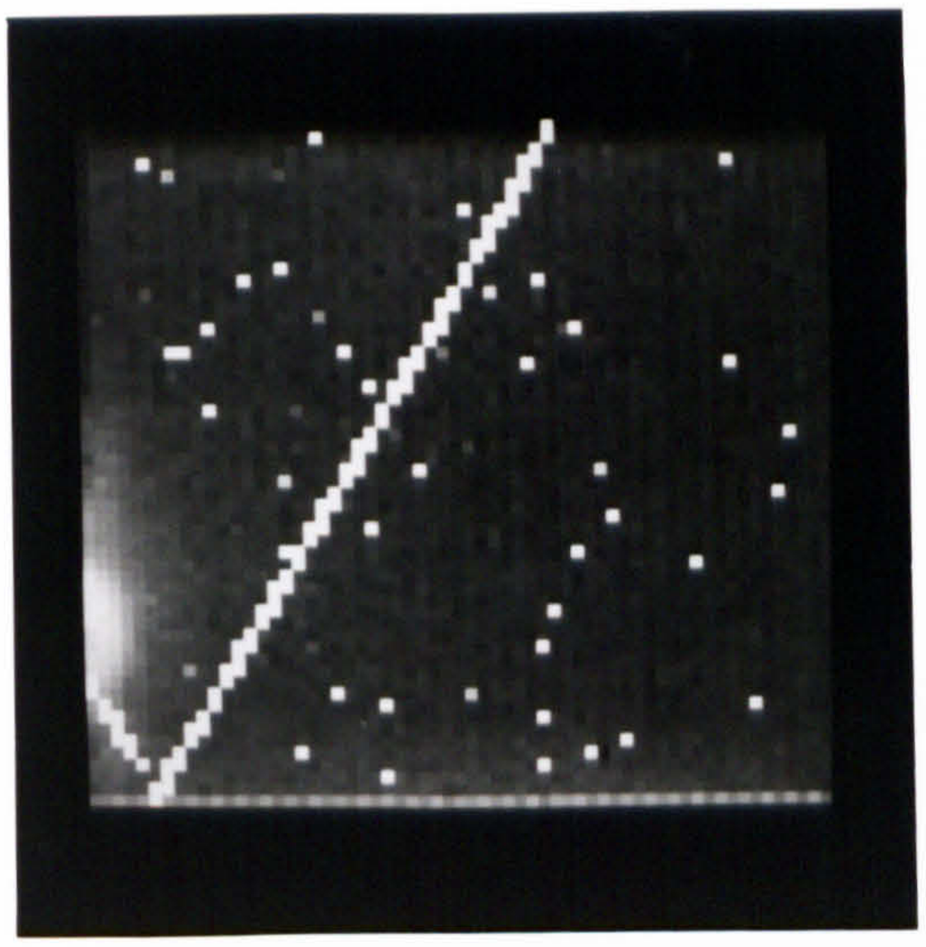


Figure 5.2: Raw dark frame

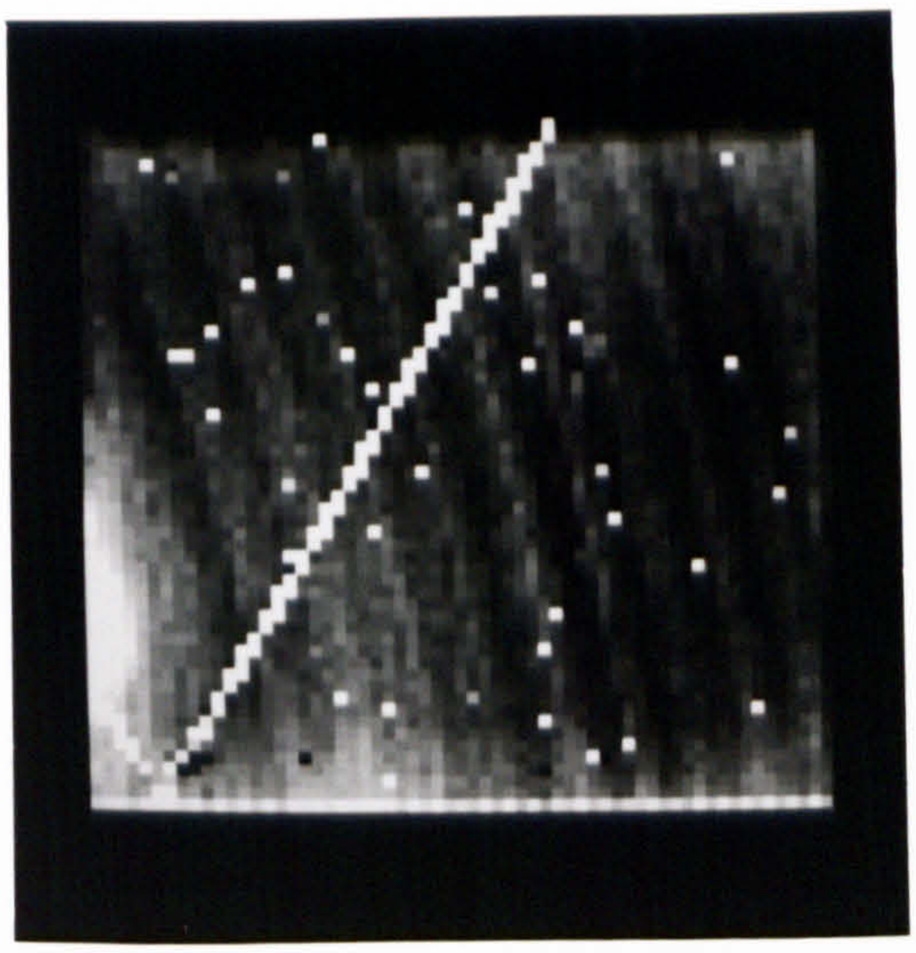


Figure 5.3: Raw sky frame

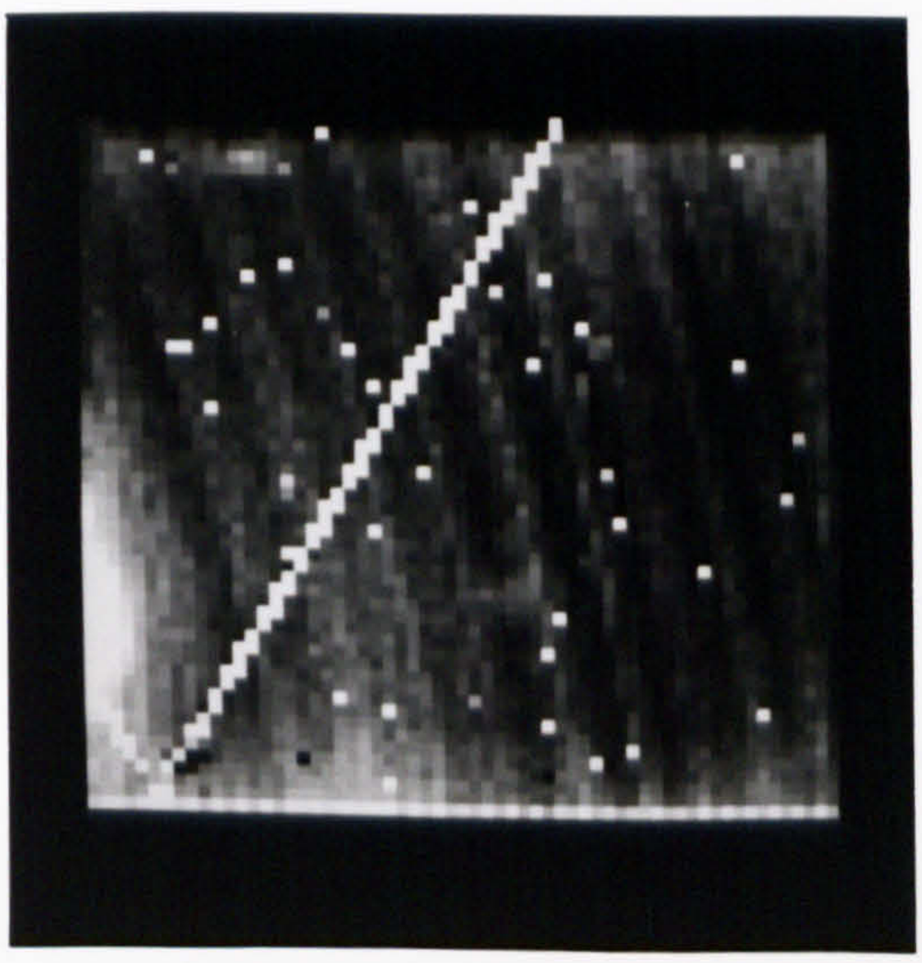


Figure 5.4: Raw source frame

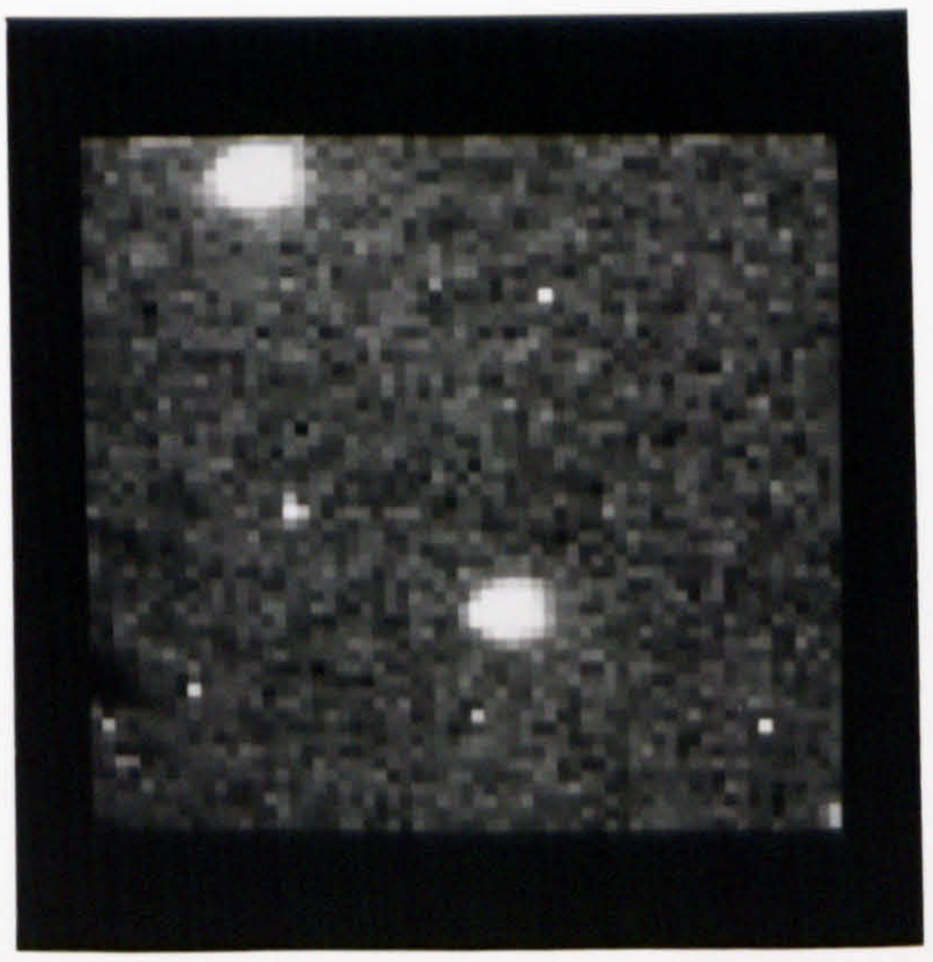


Figure 5.5: Raw source minus raw sky

the dark frame. The clear banding in the former images is due to quantum efficiency variations in the bulk detector material, due to the mechanical polishing process used to thin the backside of the detector array. The banding is seen in the illuminated images as the photons are detected in the first few microns of the detector material, which is altered during the thinning process. No banding is seen in the dark frame as the dark current arises either in the p-n junction depletion region or near the front surface of the detector, both well removed from the effects of thinning. The hotspot and bad pixels seen in the dark frame are also seen in these frames, although a few additional black pixels are seen, which appear not to detect photons despite registering dark current. The nature of these dead pixels is not fully known, but are likely to have dead MOSFETs. The only slight difference between the sky and source frames is the faint presence of two stars in the source frame, near the top left corner, and just below centre. The fact that it is very difficult to see sources in raw frames is due to the high noise level in the raw data. This noise is not random, but is due to the slight MOSFET gain variations from pixel to pixel. This constitutes a fixed noise pattern which is completely removed in the subsequent reduction. The noise level seen in these raw frames is reduced by a factor of 30 in the following reduction process. What is implied is that raw source frames are of very little use, and must be reduced in some way before they can be used even to check that the correct object is being imaged. This is in contrast to optical CCDs, where the same output amplifier is used for every pixel, and raw frames are usually very uniform.

Thus 126 source frames, 21 sky frames, and 5 dark current measurements were obtained, for a total of 152 frames. Each frame had an total integration time of 30 seconds, and thus 76 minutes of on-chip integration was obtained in three hours of telescope time, an observing efficiency of over 40%. This efficiency was lower than possible, as the telescope offset positions were typed in manually, and considerable overheads were incurred in typing commands to control the storage of data from the camera. When the telescope and camera software are fully interfaced, highly efficient observing can be expected for projects such as this, with predefined imaging positions.

5.3 Data reduction techniques

As discussed elsewhere, one of the aims of this thesis has been to provide a convenient calibration procedure which accurately removes the inherent non-linearity of the SBRC array. At the time the data presented here were reduced, this aim had not been achieved, and these data were reduced using an interim technique, namely that discussed in chapter 3, which assumes linear detectors. This technique is summarised here as :

- Subtract a dark current frame from both the source and sky frame, removing both the dark current and any electronic offsets
- Divide the source frame by the sky frame to flat field the source frame, removing pixel-to-pixel response variations
- Multiply the source frame by the median value found in the sky frame, to re-normalise with respect to the flat field
- Subtract the median of the sky frame from the source frame, to sky subtract

We shall see later that the mean sky background flux observed during the experiment was approximately 25 times higher than the K sky background predicted by SIRCAM, due to inadequate thermal baffling of the camera system. Although this extra background flux lowered the sensitivity of the camera due added shot noise, it also reduced the non-linearity to less than 1% for all sources fainter than $\sim 8^{m}.5$. As the stars used for photometric calibration were roughly of this brightness, we consider the non-linearity negligible with regards the work presented in this chapter.

Before proceeding with the outlined technique, it is worth examining the result of a far more simple process that is used as an on-line diagnostic. This involves subtracting the raw sky frame from the raw source frame. The result of doing so for our example data is shown in figure 5.5. This simple technique gives visually very good results, as it removes the electronic offset, dark current, sky flux, and most of the bad pixels. Indeed, the residual background noise in the subtracted image is virtually identical to that obtained with the more complex process. However, this technique is fundamentally wrong, and should not be used for accurate reduction. The pixel to pixel variations

across the array are due to changes in quantum efficiency and MOSFET gain. These are both multiplicative factors. Although it can be shown that a simple subtraction of the sky frame will accurately remove their variations from regions of the source frame containing the same sky level only, a simple subtraction *will not* compensate for variations in these gains for signal levels higher than the sky flux. To accurately correct for these gain variations, the source frame *must* be divided by a flat field frame, after any dark current and electronic offsets have been subtracted from both.

In general, routines from the Rapi2d package (written by this author) were used to reduce and analyse the data. A full list of these routines is given in Appendix B. For some special cases, particularly the ARGS image blinking process discussed below, hardware specific routines from the Starlink ASPIC image processing package were also used.

Unless otherwise stated, statistics are given in the following sections are derived from a 15 × 15 pixel box on the array, starting at pixel 43,39, known to be free of bad pixels.

5.3..1 Dark current frames

Prior to commencing this experiment, the detector was being run at a temperature of about 40 K in order to minimise the dark current. However, due to a sharp decrease in quantum efficiency at lower temperatures discovered in FPA004, and as seen for other early engineering arrays [42], the detector was warmed a few hours before starting to take these data, stabilising at just below 50 K. Due to thermal settling, it was anticipated that there might be some fluctuations in the dark current during the observations, so dark current frames were obtained every forty minutes. Assuming the electronic offset to be constant, a very short exposure time (130 msecs) frame was subtracted from each dark current frame, the median signal measured, and then converted to a mean dark current by dividing by the total effective integration time per frame (30 seconds) and multiplying by the electronic gain (40 e^-/DN). The results are shown in table 5.2. Note that, despite worries about dark current stability, the measured value was constant around 33,670 e^-/sec/pixel (5.4 fA) to within ±1% over the whole three hour period. Note also however, that the expected dark current at 50 K for these detectors under the

Time (HST)	Signal (DN)	I_D (e^-/sec/pix)
0145	25212	33616
0230	25176	33568
0307	25206	33608
0349	25386	33848
0438	25275	33700

Table 5.2: Variation in dark current whilst making Orion K mosaic

given operating conditions, should be as low as a few hundred electrons per second per pixel (see figure 3.7). From these and other data, it was discovered that the detector array was itself emitting photons, due to an electro-luminescent effect at one of the output FETs on the multiplexer chip. This photon flux was then detected through the front surface of the array, creating an excess dark current. We can see that this excess dark current cannot have been due to a photon leak in the camera optics, as the raw dark frames (figure 5.2) do not show the quantum efficiency variations seen in the raw source and sky images (figures 5.4 and 5.3). This electro-luminescence problem, seen in optical CCDs also, has since been resolved by optimisation of the output drain voltage. In the current data set, we can do no more than treat it as if it were dark current.

5.3..2 Sky frames

Sky frames were obtained for two distinct reasons. Firstly, as the Orion Nebula is bright and extended, it might not be possible to derive the absolute sky level from within any given image, and therefore regular monitoring of the sky level was required. More importantly, sky frames are used as flat fields. As discussed elsewhere in this thesis, and illustrated later in this chapter, the very best flat fielding is given when the sky frames are taken close in time to the source frame, even when the changes in absolute sky flux are small. As a compromise, we took sky frames after every sixth source.

The experiment took three hours, and during this time the airmass increased from 1.14 to 2.12, and we would expect quite substantial changes in the K sky background, as approximately one third of the background through that filter should be due to non-thermal OH* emission, which varies according to the van Rhijn function (see sec-

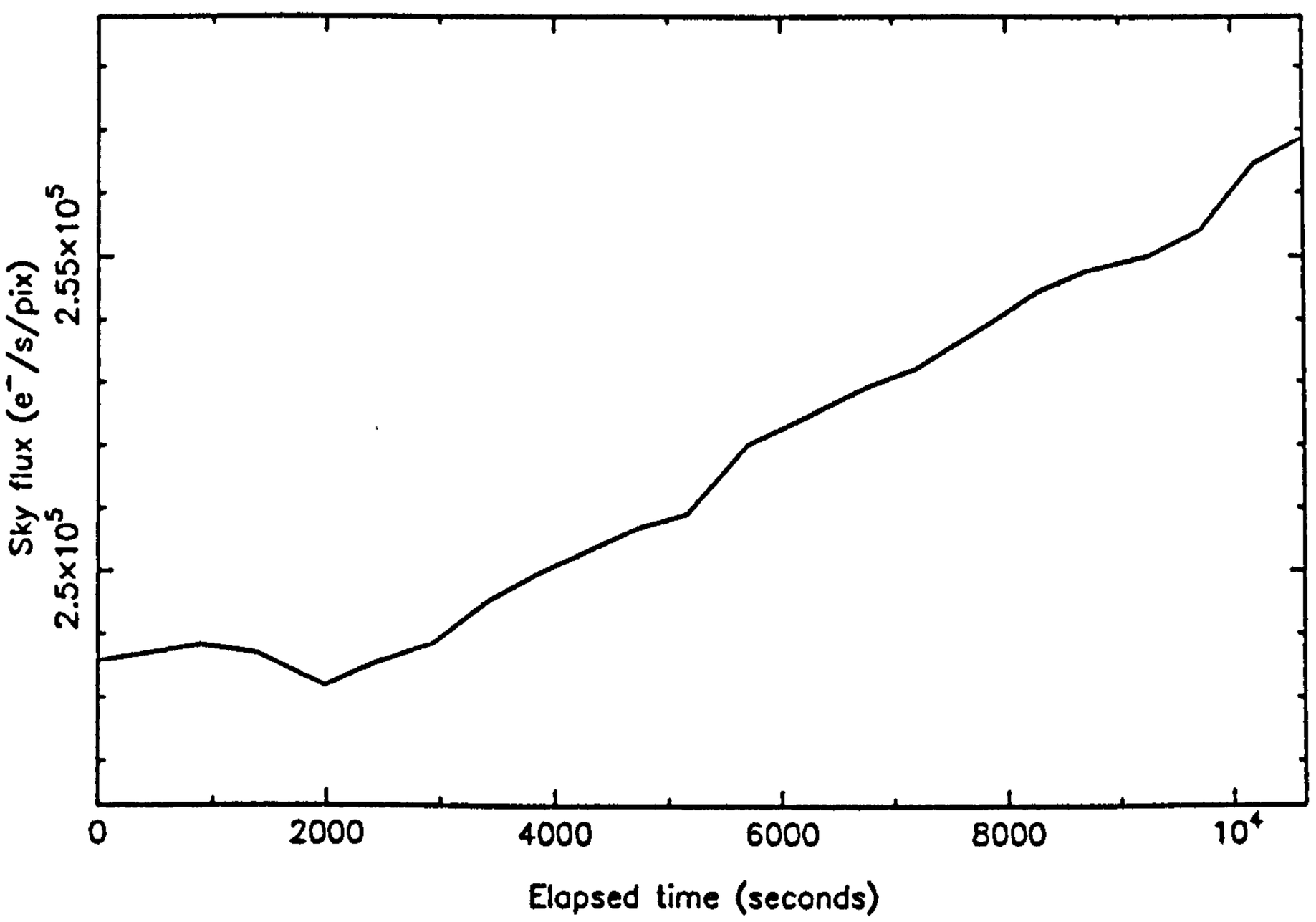


Figure 5.6: Variation in sky flux whilst making Orion K mosaic

tion 4.5.5.1), roughly linearly with the airmass.

Subtracting the nearest dark frame in each case, to remove the electronic offset and dark current, the median detected sky flux was calculated for each of the 21 sky frames obtained. The results are plotted in figure 5.6. Although there is a significant increase in the detected sky flux throughout the night, it should be noted that the y-axis in figure 5.6 does not start at zero, and that the variation is only about 3% of the total. Also, the detected sky flux is about 25 times higher than predicted by SIRCAM for this mode, and even taking into account the extreme variability of the OH* emission, it is obvious that excess background flux is reaching the detector. We can assess the nature of this excess as follows. By assuming that the non-thermal sky emission has a linear dependence on the airmass, and that thermal emission from the warm optics is independent of airmass, we can calculate from the first and last sky flux measurements that the non-thermal emission accounts for less than 4% of the total observed background. SIRCAM predictions for the background seen at K in this configuration suggest that the non-thermal sky emission and telescope thermal emission should be roughly equal, and thus it would appear that there was an extra source of thermal flux in the beam. It was later discovered that the filter wheels were incorrectly

adjusted, and that a significant amount of longer wavelength flux was bypassing the K filter, and reaching the detector. This problem has since been resolved, and the sky value at K has been measured to be within a factor of 4 of the SIRCAM prediction (McLean, private communication). The remaining error is thought to be due to a slight wavelength shift in the actual K filter used in IRCAM compared to the filter profile used by SIRCAM — as the K filter is at the edge of the thermal cut-on, a small error in the long wavelength cut-off may result in a substantial error in the predicted flux.

As approximately 8 minutes elapsed between each sky frame, it was not expected that a subtraction of the median level in nearest sky frame would be sufficiently accurate for sky subtraction. However, in common with the technique employed by users of the IRPS DC raster scanning system and by the University of Rochester group, it was hoped that a simple linear interpolation between sky values could be used [38,69]. We shall see below that this technique was inadequate for the accurate sky subtraction required in making a large area mosaic.

Before proceeding to the flat fielding, it should be noted that 5 of the 21 sky frames were eliminated at this stage, as they showed a star encroaching into the western edge of the array. As seen below, the first two columns of the array were defined to be 'bad' for other reasons, and thus only those images where the star was seen beyond the first two columns were discarded.

5.3.1 Flat fielding

The raw source, sky, and dark frames discussed above are combined to perform the fundamental processing step of flat fielding. The simple procedure outlined above requires that a dark current frame be subtracted from both the source and sky frames. Examples of dark subtracted sky and source frames are shown in figures 5.7 and 5.8. This process removes the electronic offset and the dark current. Note that the bad pixels are all black, as they were saturated in all three frames, leaving them at zero after dark frame subtraction. Most of the dark current hotspot on the left edge of the array has been removed. Again, the source and sky frames are similar apart from the faintly visible stars in the source frame. At this stage, the noise has not been substantially reduced.

The source frame is then divided by the sky frame, and the result is multiplied by

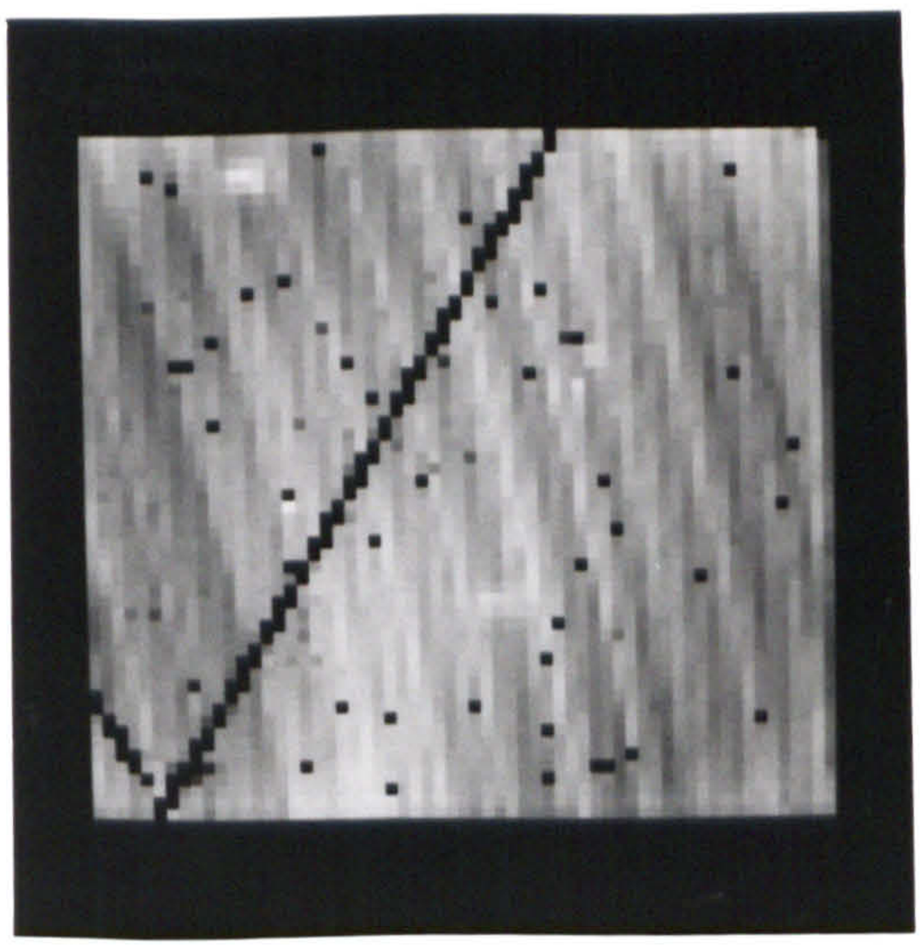


Figure 5.7: Dark subtracted source frame

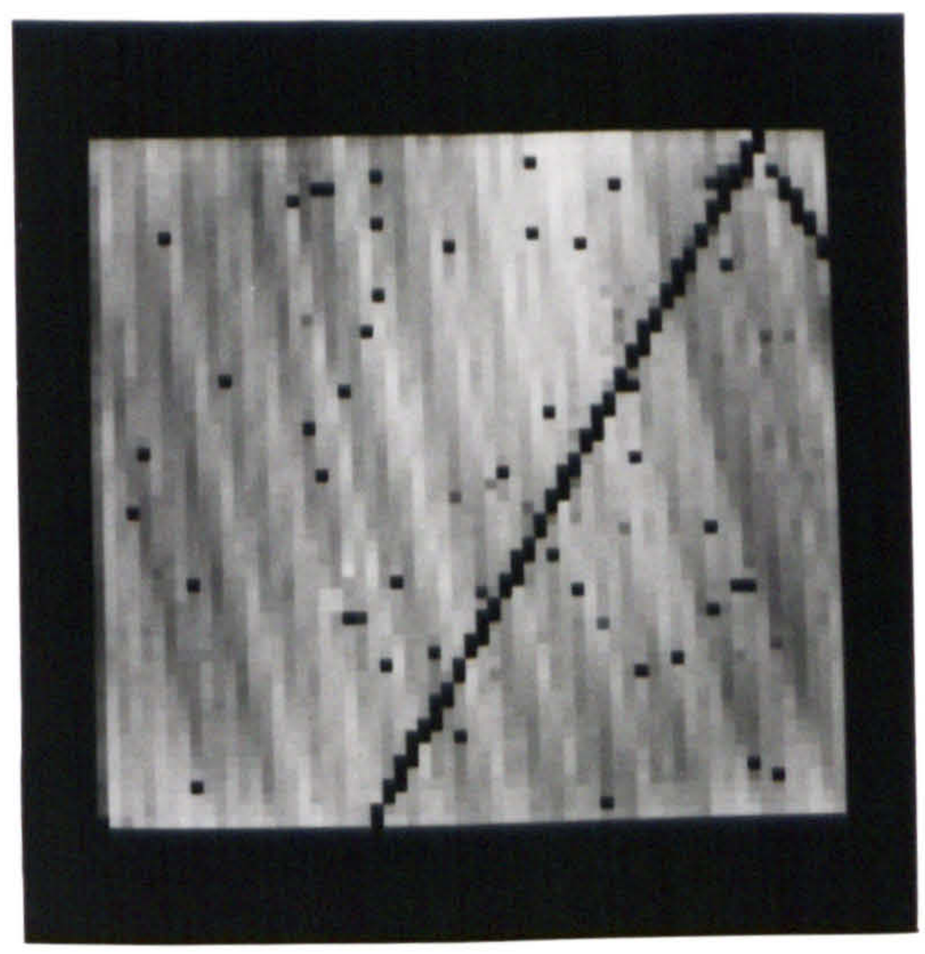


Figure 5.8: Dark subtracted sky frame

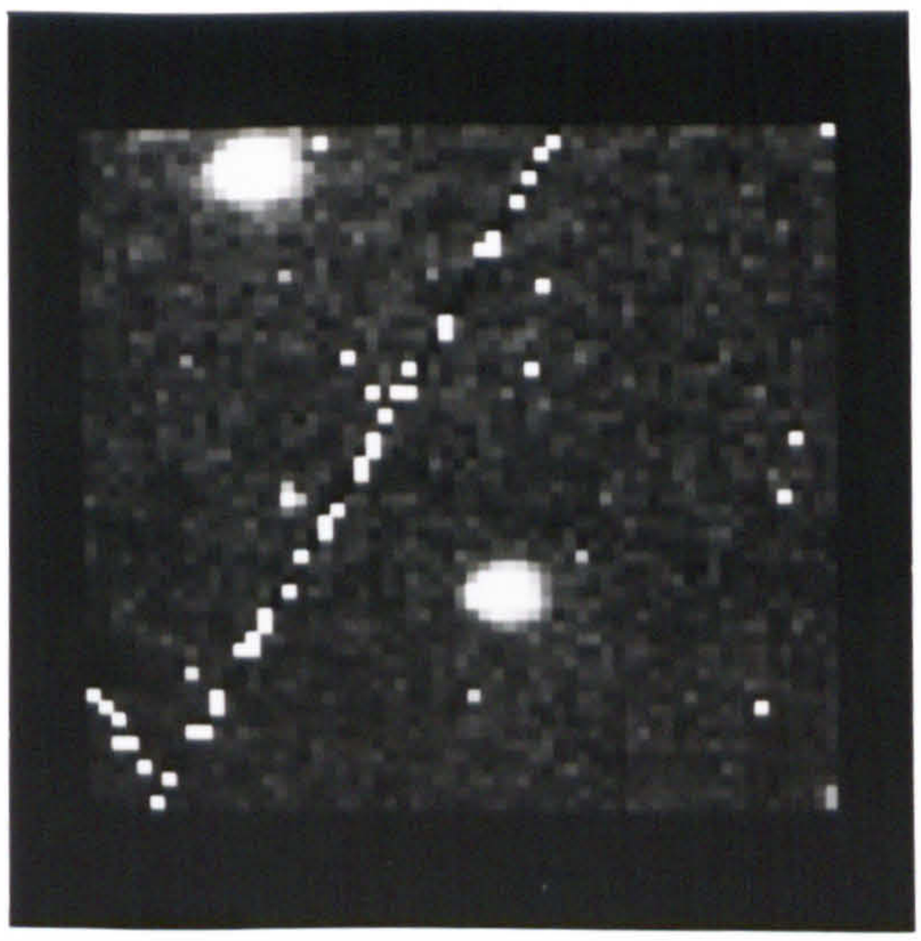


Figure 5.9: Flat fielded source frame

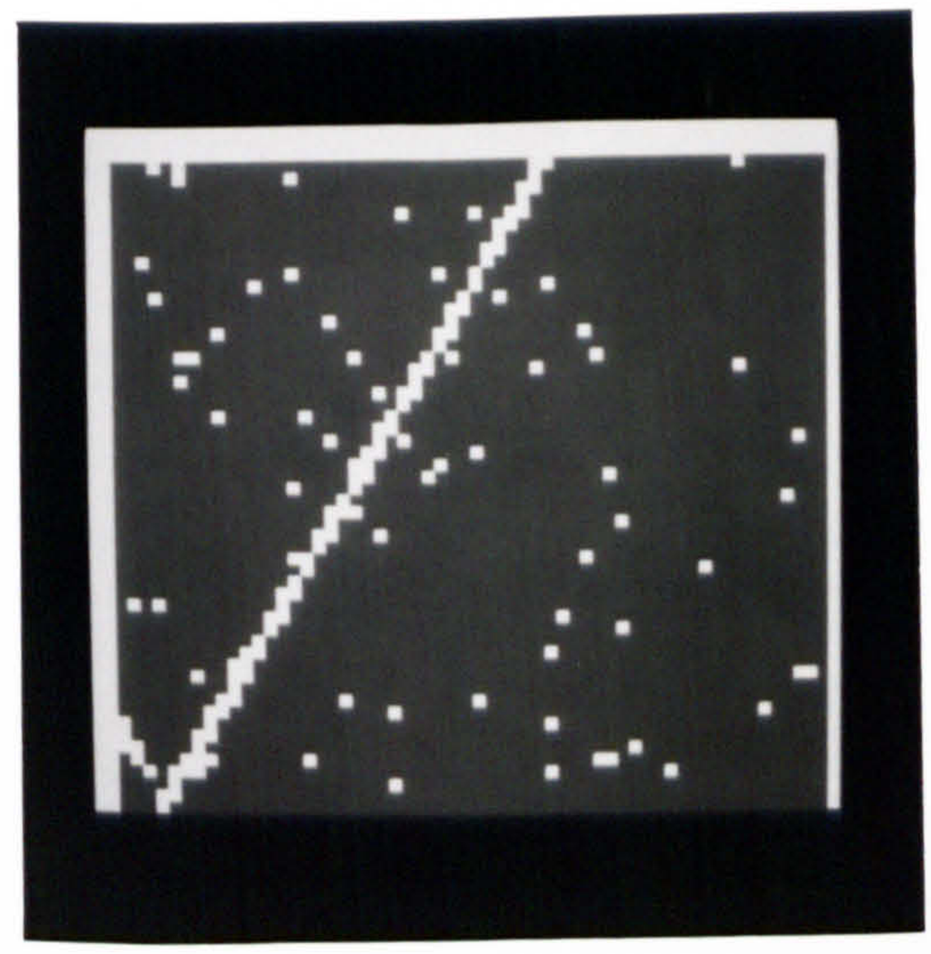


Figure 5.10: Bad pixel mask

the median value of a defined section of the dark subtracted sky frame. The result of doing this for the example frames is shown in figure 5.9.

This procedure is easily automated, and a batch processing routine (CREBAT) was used. From an input list of raw source, sky, and dark frames, the routine chooses the nearest dark frame in time to each source and sky frame, and then performs the dark subtraction for each. Similarly, the nearest sky frame in time to each source frame is used as the flat field. After the routine was run, 126 separate flat fielded and renormalised image frames resulted. At this point, no sky subtraction is performed for the individual images, for reasons discussed below.

The accuracy of the flat fielding process was estimated by examining the standard deviation over a portion of a flat fielded image clear of any stars, in the south-eastern corner of the final mosaic, beyond the bright bar. It is seen that this region is relatively free of nebulosity, and we assume that the errors would represent the true flat fielding accuracy. The standard deviation was $\sim$ 150 DN and the mean signal $\sim$ 214853 DN, which implies a flat fielding accuracy of 0.07%. Therefore, even though an engineering array was used under high dark current and high sky background conditions, the flat fielding accuracy exceeded that usually obtained with optical CCDs by a factor of 5–10. As discussed elsewhere in this thesis (see section 3.9), this high accuracy of flat fielding is not uncommon for the SBRC array.

As mentioned earlier, the highest accuracy flat fielding is achieved when the sky frame is obtained close in time to the source frame. In figure 5.11, we can see the results of flat fielding with two different sky frames, the first obtained within one minute of the image frame, and the other obtained three hours earlier. The left hand image shows the lowest noise, and was the result of flat fielding by the nearer sky frame. Even though the sky flux had changed by as little as 3%, the right hand image, showing the result of flat fielding by the earlier sky frame, is clearly noisier. It is in fact about 50% noisier than the other image.

Some experiments were performed on alternative flat fielding strategies. The simple strategy outlined is potentially improved by generating a 'master' sky flat field, with a greater signal to noise than any individual flat field. This was attempted by simply taking the mean of the 16 usable sky frames, and also by taking the median of the same images, the latter method better compensating for occasional bad pixels. In both cases, the standard deviation left after flat fielding exceeded that obtained with single frame flat fielding. Similar experiments on data obtained much more recently indicate that a reduction of residual noise by the expected factor of $\sqrt{2}$ *can* be achieved by median stacking many sky frames (Miller, private communication), and therefore we conclude that some unknown variation in the data presented here prevented our doing so, and that such variations have since been eliminated.

Finally, by inspection of several flat fielded frames, a list of bad pixels was drawn up, including dead pixels, hot pixels, and pixels which although sensitive to photons and normally behaved with regards dark current, did not flat field correctly. The latter class

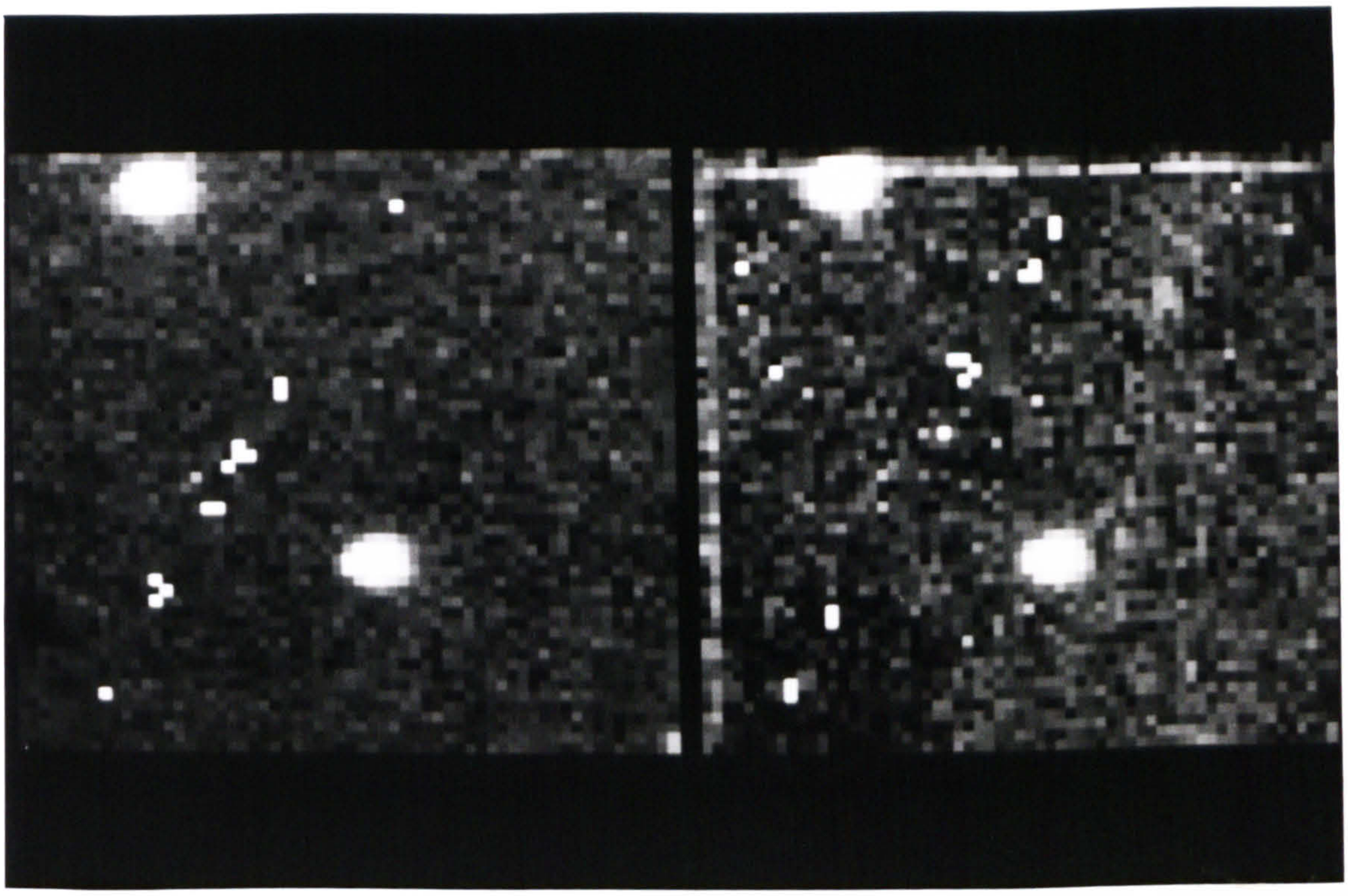


Figure 5.11: Changes in flat fielding accuracy

were all those pixels in the first two and last columns, and the top three rows, which showed up as stripes in the final mosaic unless avoided. In total, 512 pixels in array FPA004 were designated as 'bad', and a schematic representation of their positions in the array is shown in figure 5.10. The use of this bad pixel mask is discussed further in section 5.3.2.3. It should be noted that this very high fraction of bad pixels is not representative of the scientific grade arrays, which generally have only a few tens of bad pixels.

5.3.2 Image mosaicing

In the previous section, the raw data frames were reduced to 126 individual flat fielded images. The signal in these images corresponds to the source and sky flux alone. Next, we examine how these individual images were combined to produce a single continuous mosaic.

There are two basic steps. Firstly the offset of each frame from some fixed centre must be defined. Secondly, the relative intensity levels must be matched throughout the final mosaic. Both processes are simple in principle, as long as the telescope can

be accurately offset, and the sky remains constant. Unfortunately, neither of these assumptions is true for the data shown here.

5.3.2.1 Offset positions

The UKIRT was not designed as an optical telescope, and cannot be relied on to offset to the required precision (i.e. less than one pixel $\sim 0.6''$). It is also prone to long term pointing drifts, on the order of a few arcseconds. A third problem compounded the pointing errors. When doing single channel aperture photometry, the usual practice of the night assistants is to type in an offset, wait until the telescope has finished slewing, and then adjust the fine encoders until the displayed offsets match those requested. However, the offsets are only displayed to the nearest arcsecond, and manual adjustment can leave the actual position of the telescope in error by $\pm 0.5''$, i.e. ± 1 IRCAM pixel. This manual adjustment was performed as usual during the taking of these data. It was subsequently realised the problem this may create, and all mosaic imaging is now done without manual fine adjustments. However, the mechanical offset error alone is still enough to make final offset positions uncertain to greater than 1 pixel.

Thus, the offsets between the frames had to be determined from the data themselves. Starting with the central frame containing the Trapezium as the designated zero offset and spiraling outwards, the relative offset of each new frame from one previously done (and thus the cumulative offset from the Trapezium frame), was defined using one of three methods :

- Centroiding — If a sufficiently bright star was located in the region of overlap between the two images, centroiding software was used to calculate the x, y offsets between the star images. This technique is accurate to approximately ± 0.2 pixels, but cannot be used where the star is faint, in the presence of highly structured nebulosity, or where bad pixels cause the centroiding software to fail.

- Blinking — If centroiding was not possible for the reasons mentioned above, or if the only common feature to the two images was diffuse nebulosity, blinking was used. The Aspic routine ABLINK allows to images to be blinked on an ARGS screen, and one moved relative to the other until the images are offset correctly, and the common features are seen to 'lock'. The relative offset is then output by

the routine. This technique is only accurate to the nearest integer pixel, but it is usually easy to identify the correct offset at this level.

- Guessing — In the Orion Nebula, the large density of stars and considerable nebulosity meant that only very few frames were completely devoid of some feature in common with one of the surrounding overlap frames. Only five images in the mosaic fell into this category. For each, offset positions were guessed at from the mean relative offset of frame pairs with similar orientation. That is, the offset of the blank frame from a frame to its south-east (for example) would be evaluated from a number of pairs of frames with known offsets in the south-easterly direction. Although this technique is error prone, it should be noted that it has no effect on the overall accuracy of the mosaic, as by definition, the field is blank, and contains no sources that will be wrongly positioned. Nor can that frame be used to define subsequent offsets, for the same reason, and thus any error will not propagate.

Two thirds of the relative image offsets were defined with the centroiding technique, one third using the blinking technique, and five by educated guesswork. The centroiding technique should allow a final image mosaic to be made with each individual image placed to an accuracy of a few tenths of an arcsecond, by resampling the images before making the mosaic. However, it was decided to use integer offsets alone, for the following reasons :

- the sub-arcsecond accuracy of the centroided image offsets would be nullified by the integer pixel accuracy of the blink offsets
- the effects of the large number of bad pixels (see below) would be significantly increased, due to the interpolative nature of the resampling process
- the resampling process would lower the apparent spatial resolution by effectively smoothing the images

Using an integer pixel offset scheme only, the accuracy of each offset is ± 0.5 pixels, both in x and y. As each subsequent offset is defined, the error should increase as a random walk, as the square root of the number of steps. As 8 steps are required to get from the centre of the mosaic to the edge, we might expect a cumulative error of approximately ± 1.4 pixels in x and y, corresponding to $\pm 0.9''$ in both RA and declination. Combining

these errors, we can give an approximate limit to the accuracy of positions in the outermost corners of the mosaic relative to sources at the centre of $\pm 1.25''$. The error for other positions within the mosaic should be less than this. We shall see in section 5.4.3 that this accuracy limit is consistent with the astrometric accuracy derived for the image.

5.3.2.2 Intensity matching

When intensity matching images of the same source taken at different times, two factors need to be normalised before combining them. The first is the additive sky background, and the second is the multiplicative atmospheric transmission. Changes in the sky background flux can be treated as an offset between the two frames, and changes in atmospheric transmission can be treated as a gain term for the source flux.

In order to solve a similar problem for optical CCDs, we have previously developed the following technique. By accurately defining the spatial offsets between two images of the same source obtained on separate nights, it is then possible to plot the signal observed for a given position on the source on one night, against the signal observed for the same position on the source on the other night. If the detector is linear, we can make a least squares fit to the plot to derive a line whose offset on the y-axis is the difference in sky background, and whose slope is the ratio of the atmospheric transmissions. Suitable application of these numbers to one of the images will normalise it with respect to the other, allowing them to be accurately combined. Naturally, this technique assumes that the extra-atmospheric source flux remains constant.

We had hoped to apply this technique to the current infrared data, but some experiments with this least squares fitting algorithm were found to be completely unsatisfactory for the following reasons :

- To provide an accurate fit to the line, a large number of pixels in the spatial overlap region is required. The maximum overlap of any two frames in the K mosaic was a quarter of the array ($\sim$ 1000 pixels), and many of these pixels were bad in one or other of the frames.
- Also required for an accurate fit is a large dynamic range in intensity. If only sky

pixels are matched, then the scatter in plot is large, and the fit poor. To obtain a large dynamic range, stars or bright nebulosity must be included in the overlap region. However, due to the large changes in intensity that occur for small relative offsets on stars, the images must be resampled to a common centre to less than 0.2 pixels for the technique to work. As discussed above, resampling tends to propagate the effects of bad pixels, and was to be avoided for these data.

- The data provided by infrared arrays frequently has a lower dynamic range than from CCDs due to the bright sky, and the higher, read noise and sky shot noise tend to increase the scatter of the plot still further.

Thus, the only intensity corrections applied to the data were an additive offset correction, due to changes in the sky brightness. It is these offsets which would be most visible in the finished mosaic as a 'patchwork quilt' effect.

Not correcting for the changes in the atmospheric extinction with airmass will not be as visible in the final mosaic, but will cause an error in the relative brightnesses of sources across the array. As no standard stars were measured whilst these data were obtained, the actual value of the atmospheric extinction for the given night is unknown, and no corrections for atmospheric extinction have been made in these data. However, assuming a median K extinction for Mauna Kea of $0^m.07$/airmass [83], and noting that the typical difference in airmass from between pairs of offset frames is ~ 0.2, we can see that the error in the relative intensity of a star common to both frames is on the order of 1.5%. The total airmass change between the first and last frames of the mosaic was ~ 1, resulting in an error of approximately 7%.

The method used to derive the sky intensity offset was simple. The median value was calculated for a box containing mainly sky in the overlap region of a pair of frames, and the difference in these values was subtracted from the relevant frame. The result of this process was checked by making a small mosaic of the pair, and checking for any residual 'patchwork' effect at the boundaries between the images and the overlap region. This visual checking was very valuable, as the eye is very sensitive to faint linear features in the presence of noise. On several occasions, ghost images and reflections in the statistics boxes caused the intensity matching to be poor, and iterating on the sky correction process resulted smoothly continuous image we shall see below. Examples of a pair of offset frames are shown in figures 5.12 and 5.13, where both image pairs have

been offset according to the relative positions of the common star, but where the first image shows the result obtained with no sky intensity matching, and the second shows the result after matching. Note that the bad pixels in the images have been masked out, and are represented as black. Note also that in the intensity matched pair (figure 5.13), the noise in the overlap region is noticeably less than in the regions outside the overlap. On examination, the reduction of noise in this region is by a factor of $\sqrt{2}$, as expected.

This stage of the processing was the most time consuming of the whole reduction procedure, due to the interactive nature of the methods used to define the position offsets and sky intensity offsets. After the spatial and intensity offsets had been defined for the 126 source frames, the composite image mosaic was made.

5.3.2.3 Making the mosaic

This next step is in principle simple. All the input images are read in, along with their offsets from some defined central point. From the sizes of the input images, and the minimum and maximum offsets in all directions from the central point, the size of an output image large enough to cover the whole imaged region is calculated. An output image is then created, where the value of each pixel within its boundaries is defined as some function of the values of all those pixels in the input images which measured the corresponding point on the sky.

In practice, this problem is more complex. Firstly, what combination of the input pixel values should be used to derive an output value? For a small area of sky imaged repeatedly, with small offsets between each image, each output pixel will have been imaged many times, by many different input pixels, and the median of all the individual measurements should be used. However, for this work, we have examined a large area of the sky, imaging each point only twice. In this case, the median becomes equal to the mean. For this reason, and for reasons explained below, we have used the mean in this work.

The major problem involves bad pixel handling. As we have seen earlier, 512 of the 3596 pixels of the detector array used in the current experiment had been designated 'bad' (see figure 5.10), and we need a scheme that minimises their effect in the final mosaic. We use a system of 'bad pixel masking' to achieve this as follows. At the same

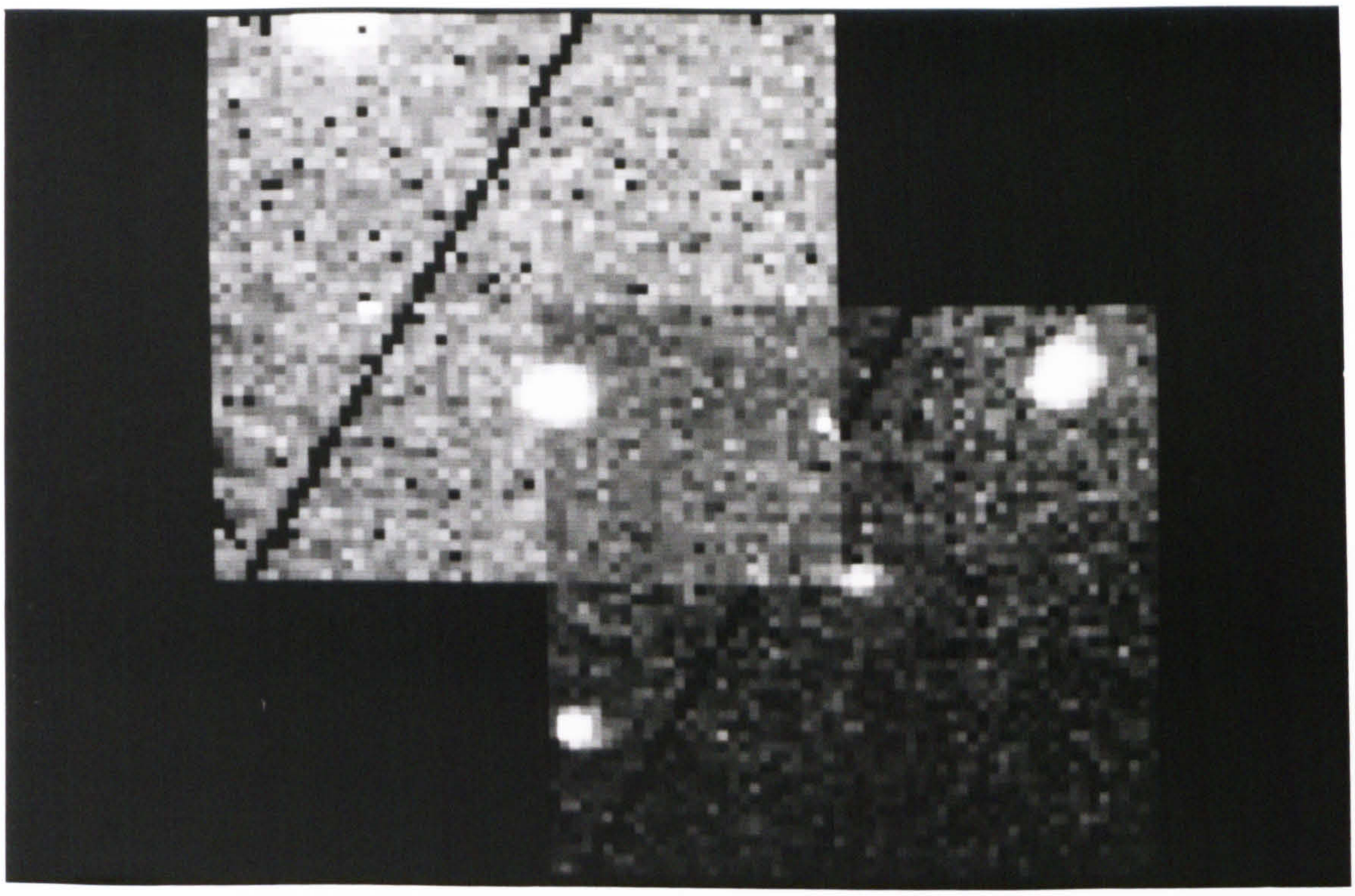


Figure 5.12: Mosaic pair with no sky intensity correction

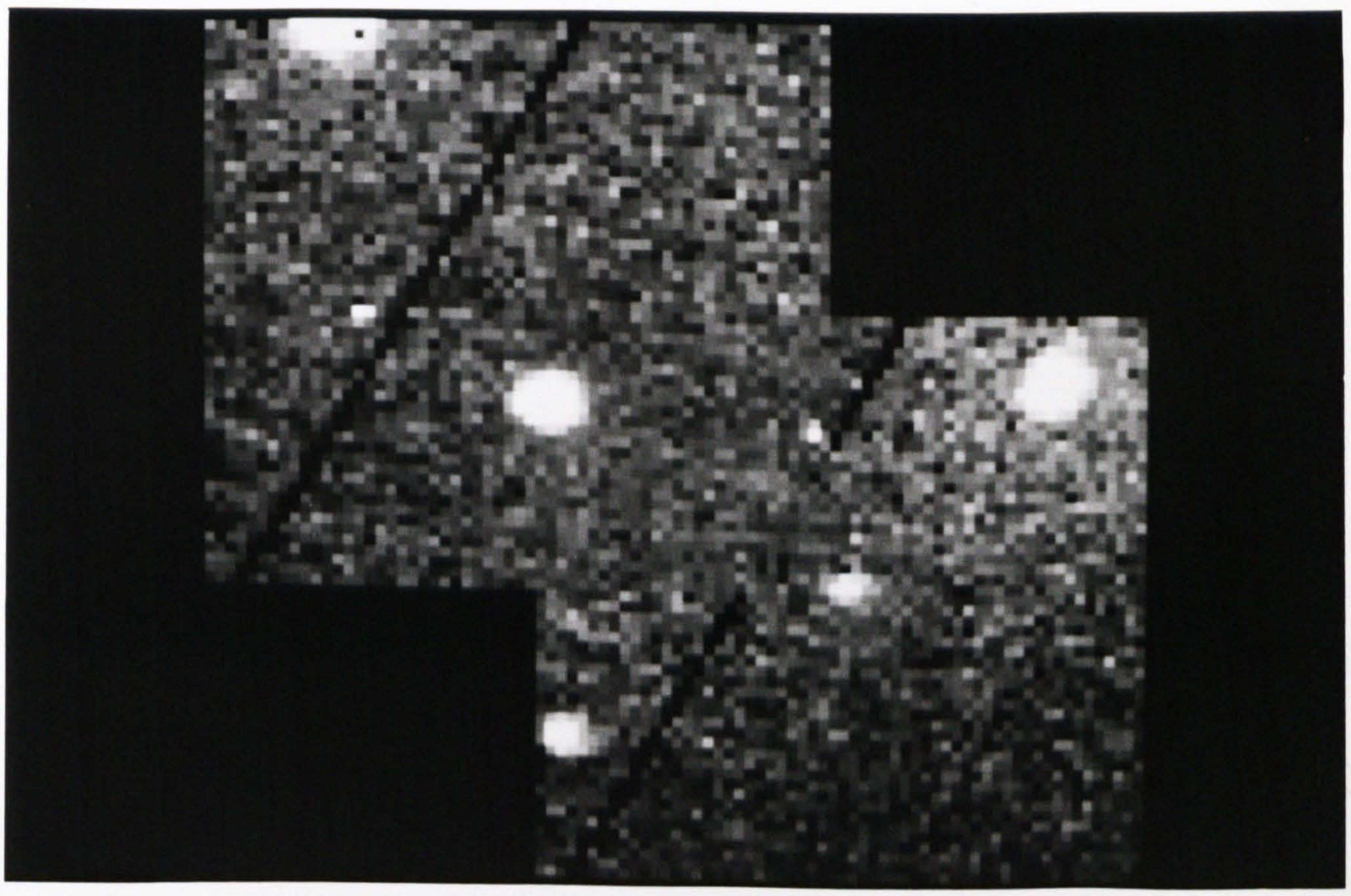


Figure 5.13: Mosaic pair after matching of sky intensities

time as reading in all the input images, we also read in a 'bad mask', which defines the locations of the fixed bad pixels within the detector array. When we create the output image, we also create a temporary reference array of the same size. We proceed thus : we take each input image in turn, calculate where it lies within the output image, and then add its pixel values one by one to the corresponding pixel in the output array. If a valid input pixel value is added to an output pixel, a record of this fact is kept in the reference array by adding 1 to the corresponding pixel. If however an input pixel is marked as bad in the 'bad mask', its value is not added to the output image, and the reference array is not updated. This procedure is repeated until all the input images have been processed.

At the end of the sequence, each pixel in the output array contains the total of all the valid input image pixel values that corresponded to that point on the sky, whilst the reference array contains a note of how many such valid additions took place for each pixel in the output array. Then, the output array is divided by the reference array, in order to normalise the intensities seen in each output pixel. If a given output array pixel has had no valid input image values added to it, then the output pixel is given a special value which flags it as being a residual bad pixel. These are then dealt with further, as discussed below. As the output image is rectangular, whilst the actual mosaic may be non-rectangular, some pixels in the output image may lie outside the boundaries of the actual area imaged, and these pixels are given another special value which flags them as being null.

In this way, bad pixels in one image are replaced with valid values from the overlap images in order to reduce the number of bad pixels seen in the final mosaic. The possibility of 'double hitting', the chance that different randomly distributed bad pixels lie at the same position on the sky in two images, is easily calculated as the square of the original fraction of bad pixels. For example, FPA004 has been seen to have a bad pixel fraction of 14%, and we might expect the fraction of bad pixels contained in the final mosaic to be about 2%.

It should be noted that this technique is very effective for limited numbers of isolated bad pixels. If large contiguous regions of the detector array are designated as bad, then they may still be replaced by data from other images, where the given region has fallen on a good part of the detector. The 'patching' might not be invisible however, as the

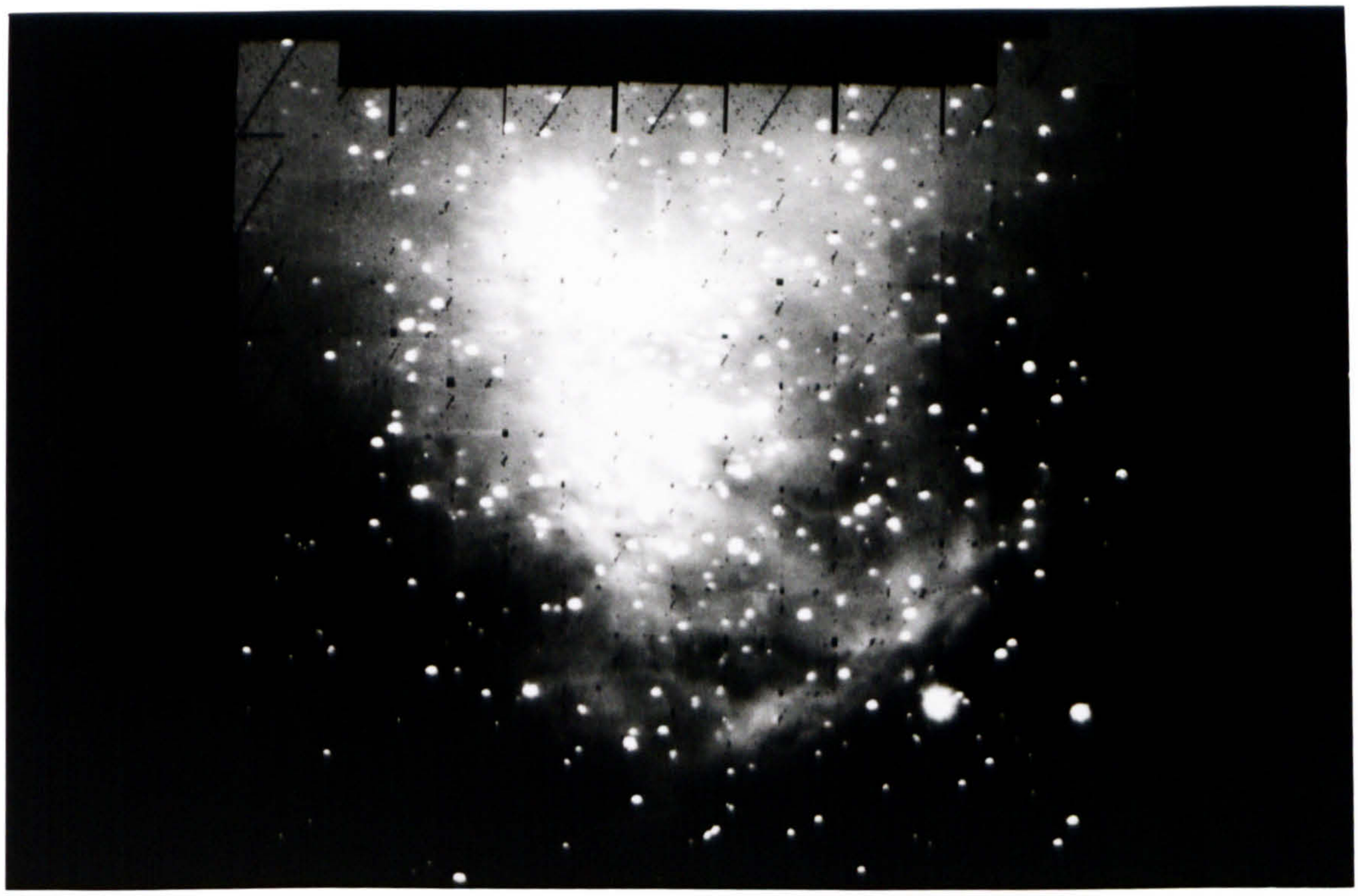


Figure 5.14: Raw K mosaic

noise is is noticeably reduced in regions where the values of two or more valid pixels have been combined, compared to regions made up from the data of one image alone. This was illustrated in figure 5.13.

A routine called QUILT was written to perform to procedure outlined above, and the 126 individual images were processed accordingly. The resulting mosaic image is shown in figure 5.14. This is the raw mosaic, in the sense that all the residual bad pixels are visible as black dots, patches, and lines throughout the mosaic. In the next section, we discuss how this mosaic was cleaned in various ways.

5.3.3 Cleaning the mosaic

After the mosaic making stage, many of the effects of the large number of bad pixels had been removed, but the result still needed some cosmetic processing to remove remaining blemishes. There are four main types of blemish that need to be considered :

- isolated known bad pixels, resulting from 'double hitting', where the bad masking process failed to find any valid data for a given pixel in the mosaic

- periodic patches and strips of known bad pixels
- random bad pixels, due to cosmic ray events for example
- optical ghosting and other spurious reflections

The first of these were cleaned using an automatic routine (GLITCH) that searches an image for a specified bad pixel flag value, and on finding one, replaces it with the median of the surrounding eight pixels. This is a very common technique, and results in invisible mends for isolated bad pixel values that lie in fairly flat regions of an image. Less than 2000 such pixels were flagged in the inner region of the mosaic (i.e. the region excluding the edge strip that had been imaged only once), and this is $\sim 1\%$ of the pixels in this region. This fraction is smaller than the anticipated 'double hit' fraction of 2%, due to the fact that the bad pixels were not distributed randomly in the detector array, but lay mainly along the edges and the crack.

There are two main periodic patches of bad pixels. As explained in section 5.2, the telescope offset pattern was defined assuming a detector field of view of $38.75'' \times 36.25''$. However this did not account for the fact that the first two and last columns, and top three rows did not flat field correctly, and were masked out of the final mosaic. Thus, the actual field of view of the detector was closer to $37'' \times 34''$, and the adopted spiral offsets lead to occasional small regions which were never imaged by a good part of the detector. This is explained schematically in figure 5.15. Only fifteen of these small holes, each about 2×2 pixels, appear in the mosaic. The automatic deglitching routine mentioned above removed some of them, but for a few 3×3 patches, the median technique did not work, and the remaining bad pixels were removed with an interactive patching routine from Aspic (PATCH). This routine fits a surface to a defined annulus around the region to be cleaned, and then fills the region with interpolated values and simulated noise. Although subjective, this routine can be very effective if used sparingly.

The other type of periodic structure is due to exactly the same error in the telescope offsets, and occurs at the edge of the mosaic. At the edge, the small holes become strips of bad pixels. Their removal required considerably more patching with interpolated data than was true for the small holes in the centre of the mosaic, but was felt worthwhile for the following reason. The rectangular subset of the mosaic that lay within the region of the mosaic that had been fully sampled (i.e. each output pixel imaged twice)

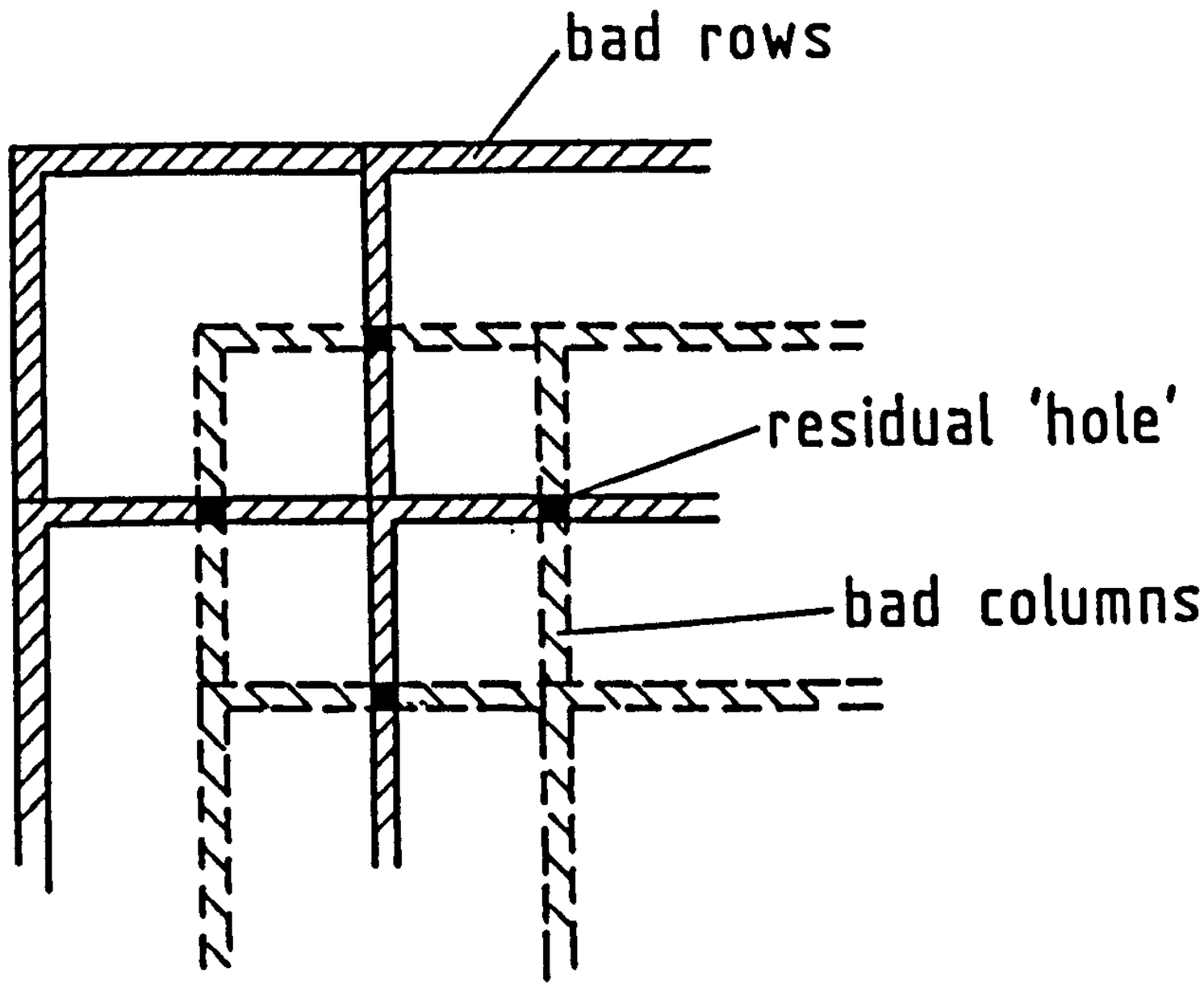


Figure 5.15: Explanation of residual holes in mosaic

was 450 x 388 pixels. By including the outermost strip of the data that had been imaged only once, the final image was increased in size to 506 x 448 pixels, a gain in area coverage of almost 30%. Although including these additional data was considered worthwhile from the point of view of spatial coverage, it was done in the knowledge that a small number of faint stars may be missed from the edges of the final mosaic if they fell on the bad pixel strips. A check of each individual image that was located at the edge of the mosaic that revealed few if any stars were missed in fact. A second caveat is that the noise along this strip is approximately 50% higher than towards the centre, due to it having been imaged only once.

Random bad pixel events, unrelated to the known fixed bad pixel pattern, are largely due to cosmic ray events. Using a routine that searches for large single pixel deviations from the local median (NITPIK), a few hundred cosmic rays events were automatically cleaned from the image. It should be noted that this type of algorithm relies on the assumption that no real image feature is pointlike. Thus it may be applied to seeing limited data such as this, where the FWHM is roughly 2.5 pixels across, but would not be applicable to data taken in the 2.4″/pixel mode, for example, where point sources may fall completely within one pixel.

The final kind of blemish is due to spurious optical reflections in the camera optics. In the raw mosaic, two long streak reflections may be identified to the left and above the brightest star in the mosaic, θ^1 Ori C. These reflections of the bright star are seen several image widths away from it, and when the star is well out of the array field of view. These reflections were removed using the same patching algorithm discussed above.

5.3.4 Making the final image

At this stage, the image is 'clean'. Three simple remaining procedures are required to produce the final image. Firstly, it can be seen that edges of the mosaic are slightly staggered on the top edge and right hand side. This is explained by the orientation of the detector array in the camera being slightly rotated with respect to the nominal north-south, east-west axes. Using the results of the astrometry discussed below in section 5.4.3, we determined the image rotation to be less than 1° east of north, and an image rotation routine (ROTATE) was used to rotate the mosaic by this small amount.

Then a rectangular subset of the data was extracted from the mosaic, the main loss being the small section of data protruding to the north-east. The final image size at this stage 506 × 448 pixels, which at the nominal scale size of 0.625″/pixel, corresponds to a total mosaic field of view of 5.27′ × 4.67′.

The final procedure necessary was to flip the image about the horizontal axis, giving us east to the left and north to the top, as is conventional for astronomical images. The resulting image is shown in figures 5.16 and 5.17, both displayed with linear scaling, at low and high contrast respectively. The image is discussed in considerable detail in the following sections, and the positions of objects referred to are marked on overlay sheets with each image. Additional images are shown throughout the text, each examining a specific portion of the mosaic.

5.3.5 Summary of data reduction

We have examined the data reduction techniques used to reduce a large number of raw frames to one final mosaic. The techniques involved are fairly simple, and will be familiar

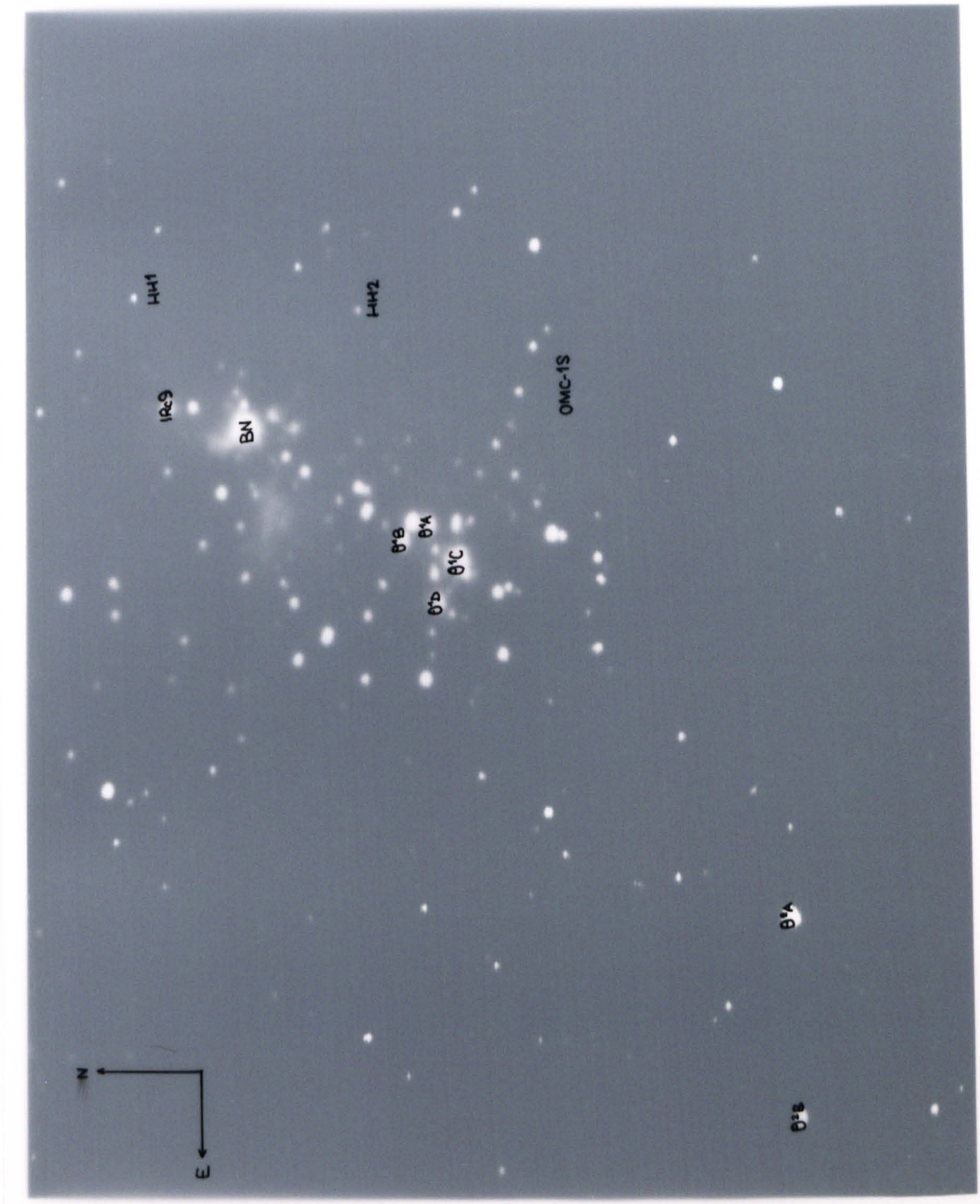


Figure 5.16: The final K mosaic of the Orion Nebula at low contrast

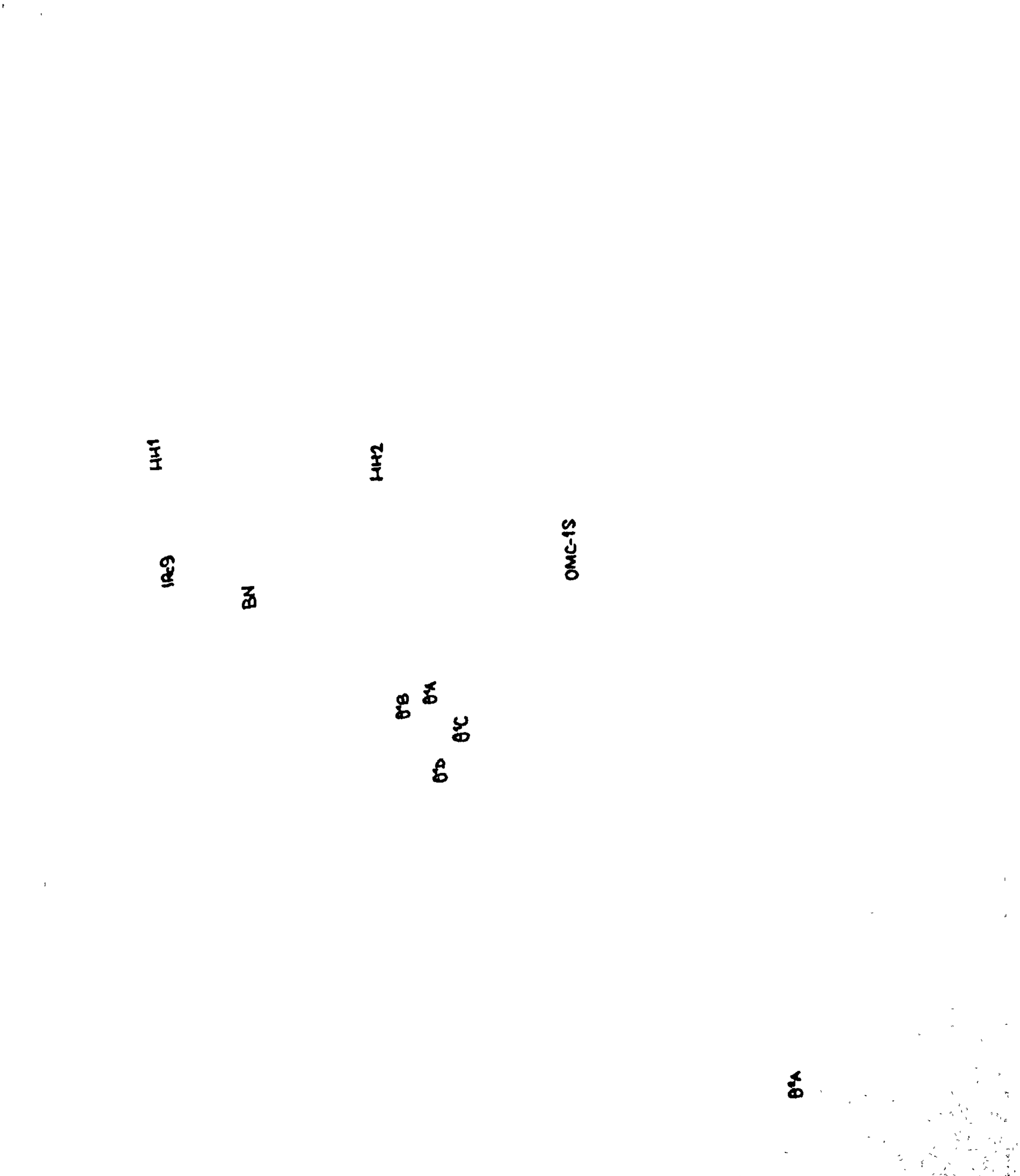


Figure 5.16: The final K mosaic of the Orion Nebula at low contrast

Figure 5.16: The final K mosaic of the Orion Nebula at low contrast

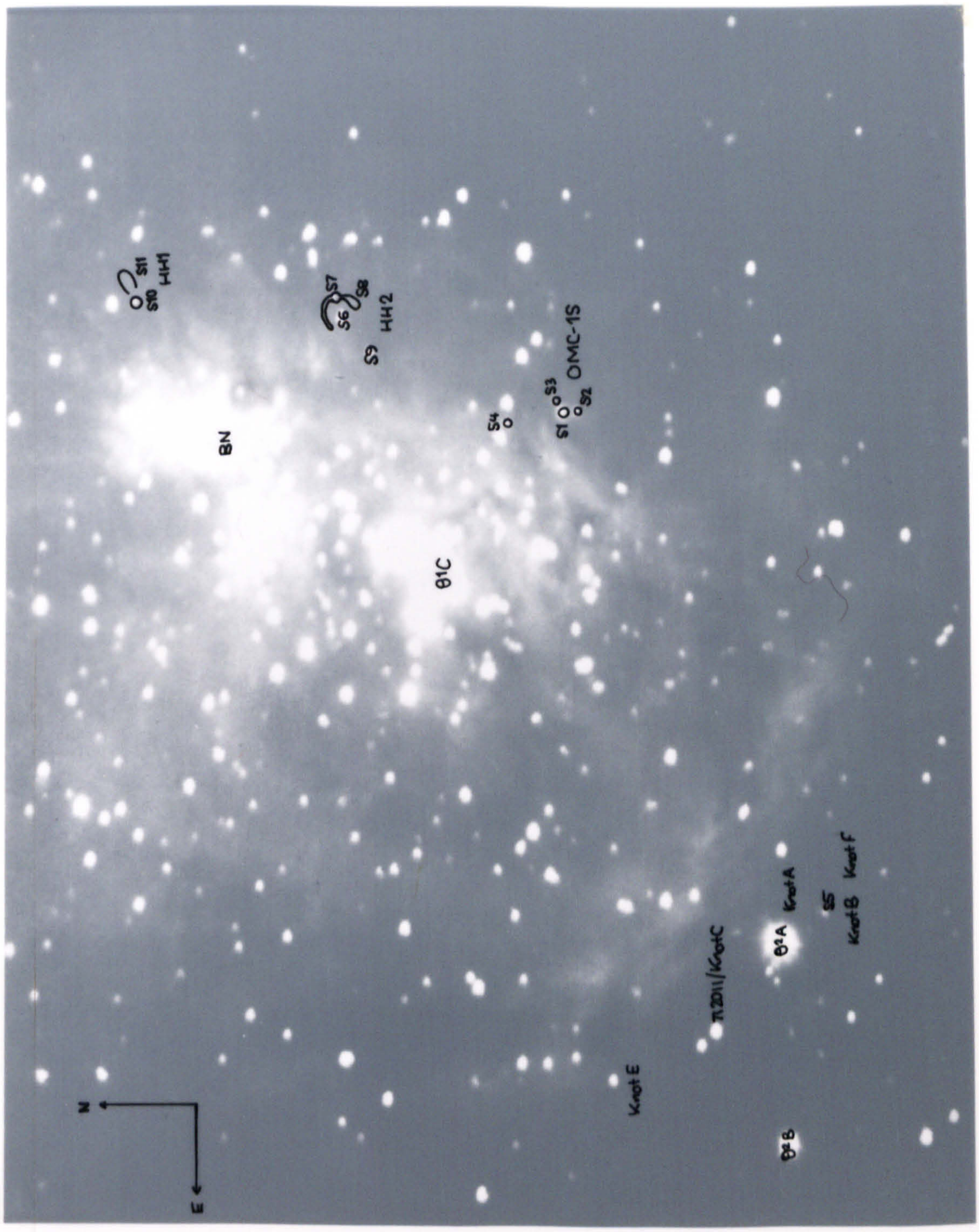


Figure 5.17: The final K mosaic of the Orion Nebula at high contrast

N

E

S10 S11

HH1

BN

S6 S7

S8

S9

HH2

θ1C

S4

S1 S3

S2 OMC-1S

Knot E

IR2011/Knot C

θ2B

θ2A Knot A

S5

Knot B Knot F

Figure 5.17: The final K mosaic of the Orion Nebula at high contrast

to optical astronomers at least. However, it is clear that infrared camera imaging is not exactly the same as optical CCD imaging, and that special account needs to be taken of flat fielding methods, frequency of sky measurement, background and extinction changes during the making of a contiguous mosaic, and for the early arrays at least, bad pixel management. In the following section we shall see how the data reduced here compare with previous studies of the same region, allowing us to assess the quality this new technique more quantitatively.

In this section, we have seen how one final mosaic image of the Orion Nebula at K was made from 126 raw source frames, 21 raw sky frames, and 5 raw dark current frames. As noted in section 5.2, only one hour of integration was required to obtain these data. However, over one hundred hours were spent in reducing these data to the level presented here, by a relatively experienced user of image processing software and techniques. Admittedly, much of this time was spent in the offset position and background intensity matching stages, and this may not be applicable to other experiments with infrared cameras. What is nevertheless clear is that the time taken to reduce data from an infrared camera may be as much as two orders of magnitude longer than the time taken to obtain it, much the same as for optical CCDs. And we have yet to discuss the actual data analysis that follows the reduction. Infrared astronomers will have to be prepared to invest considerably more effort in the reduction of their data, given the huge flood of camera data that is likely in the coming years.

5.4 Results and analysis

In this section, we shall examine the K mosaic, firstly in general terms, and then concentrating on more specific points of interest. In addition, we shall briefly discuss an H (1.6 μm) mosaic of a smaller region, and one single frame obtained at narrow band L (3.6 μm). Throughout, we shall refer to 'optical images' of the nebula, and we have included one such image for comparison purposes. It is a photograph by David Malin, taken at the AAT. In fact, it does not represent a typical optical view of the region, as it is an I-band plate taken at 8000Å, and the the shorter wavelength emission line nebulosity that usually obscures all but the very brightest Trapezium stars is well suppressed. We shall use this image as a reference in the following sections — it is shown

in figure 5.18.

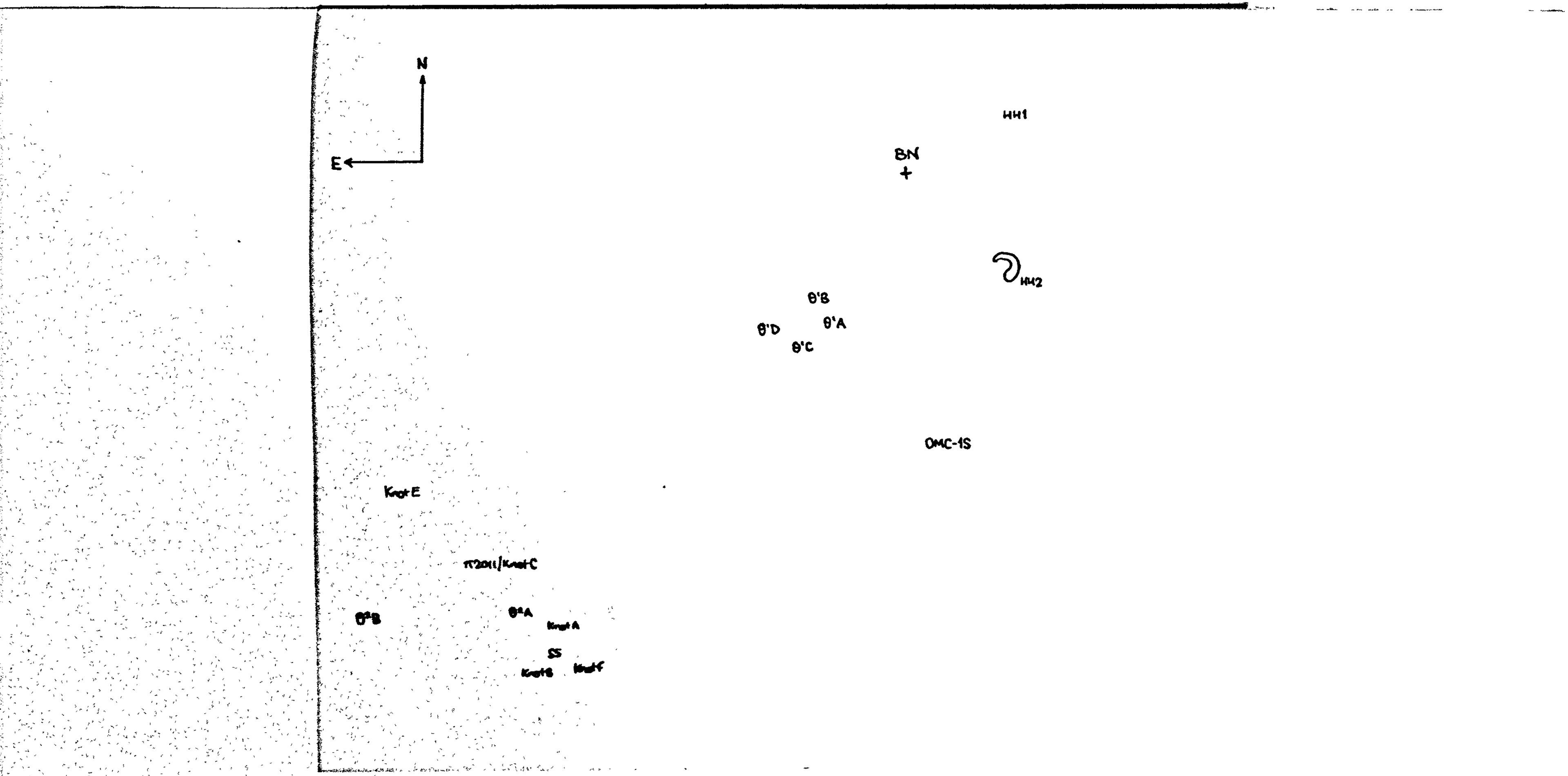
5.4.1 General examination of the K mosaic

Our new K image of Orion clearly reveals the potential of infrared imaging arrays. At last, we are able to make infrared images of obscured regions that are comparable to those obtained with optical CCDs in terms of their spatial resolution, area coverage, dynamic range, and sensitivity to highly structured diffuse emission. In fact, so remarkable is the similarity in overall quality between this image and ones made by optical CCDs, it is worth reminding ourselves that some 15–30 magnitudes of visual extinction lie in front of the infrared sources in this new image, and that infrared astronomy has clearly taken a very large step forward.

In the infrared, the Orion Nebula looks very different when compared to the usual optical views. The emission nebulosity associated with the HII region is considerably suppressed, and the very dense cluster of stars surrounding the Trapezium OB cluster is revealed. Also, much of the extinction due to dust and gas in the background molecular cloud is stripped away, and we can see beyond the HII region and into the cloud. This is particularly noticeable to the north-west of the Trapezium, where a bright infrared cluster of even younger stars is located, dominated by the BN source. Another region where the extinction has been penetrated is towards the region known as the dark bay, to the east of BN. In the optical, nothing is visible towards this dense intrusion of dust, but here at K, many faint stars are seen to lie within it or behind.

Although much of the optical extinction has been penetrated, the influence of obscuration is nevertheless still seen here in the infrared. Particularly striking is the dark 'finger' of extinction that runs up to the west of the Trapezium, ending just to the south-west of BN. Due south of BN, there is considerable extinction, and a dark band running from south-west to north-east is known to be associated with the massive disk surrounding the source just visible in the band, IRc2. From high resolution polarisation studies in the near infrared, with IRCAM and single channel photometers [11,68], it is known that the polarisation is very high and centrosymmetrically arranged about IRc2 on the south-east side of the dark lane. A similar pattern on the north-west is distorted by the addition of an additional centrosymmetric pattern, this time oriented

Figure 5.18: The Orion Nebula at 8000Å (©AAT 1983)


N
E
HH1
BN
HH2
θ'B
θ'A
θ'D
θ'C
OMC-1S
Knot E
π2011/Knot C
θ²B
θ²A
Knot A
SS
Knot G
Knot F

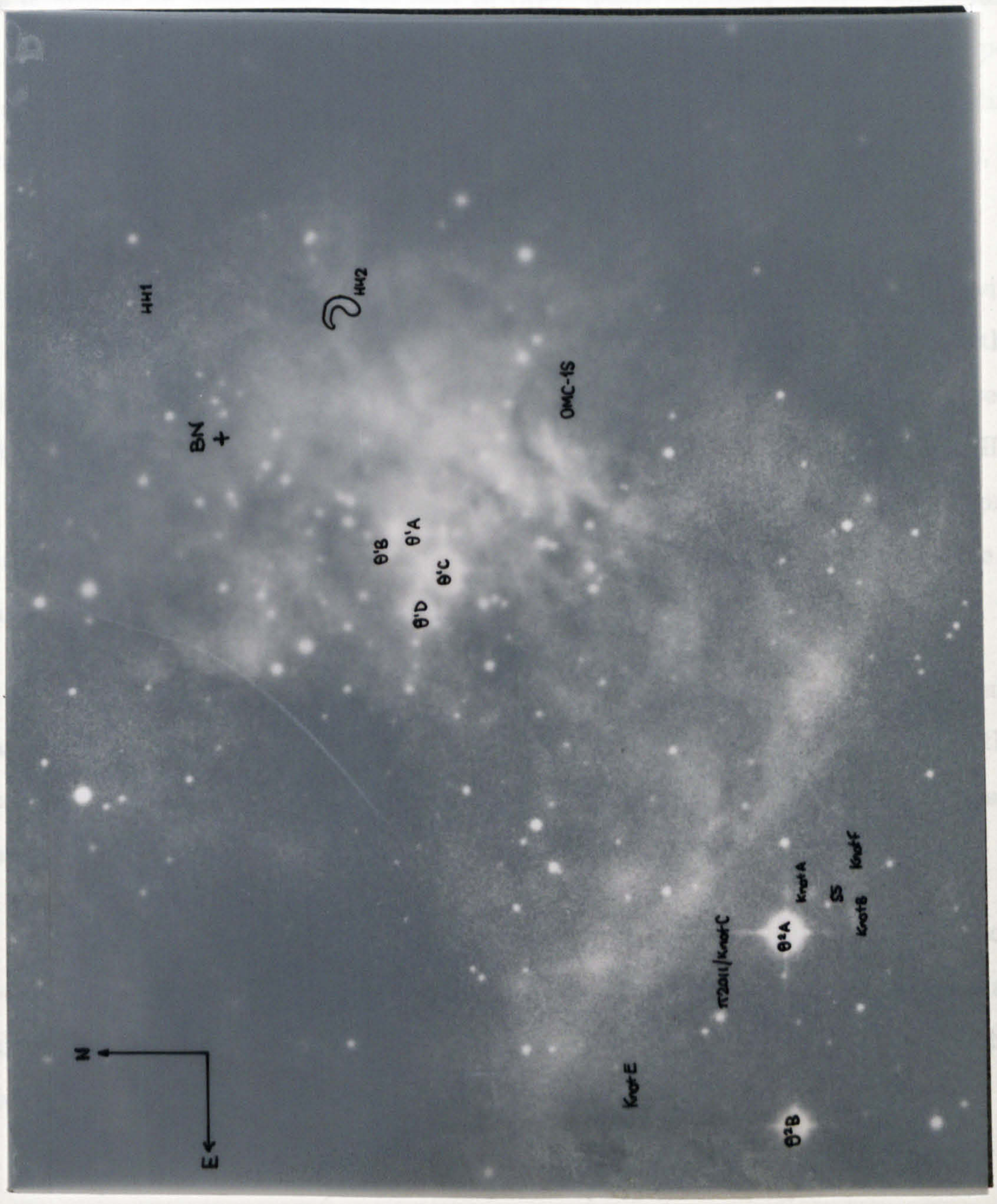


Figure 5.18: The Orion Nebula at 8000Å (©AAT 1983)

around BN. From the K mosaic alone, we can see that there is an essentially bipolar pattern centred on IRc2, with lobes to the south-east and north-west, perpendicular to the disk. The south-eastern lobe is very bright due east of IRc2 and known to be highly polarised there, but is heavily extinguished by foreground extinction to the south — the fainter nebulosity seen towards the south is not highly polarised. Several stars are seen in projection against this extinction, and are all seen in the optical. The north-eastern lobe is broken by extinction to the west of IRc2, where the dark 'finger' terminates, whilst the northern part of this lobe is broken by BN, which has its own associated polarisation pattern. In fact, polarisation due to BN appears to be less extended than that associated with IRc2, and is largely confined to the bright 'plume' to the north-east, in which the polarisation vectors indicate BN to be the illumination centre.

Even though considerably suppressed in the infrared, the nebulosity associated with the HII region to the south-east of BN-KL is still bright and highly structured in this image. Still prominent at K, the bright bar to the south-east of the Trapezium is seen to be similar in structure to the view seen in optical emission lines, particularly Hα. As the K window includes the Brγ line of ionised hydrogen, it is tempting to interpret much of the emission seen at K as due to this line, given the close correspondence in structure. We shall discuss this point further below.

Other localised patches of nebulosity are seen throughout the image, in particular near the bright star to the south-east of the bar, θ^2 Ori A. This nebulosity, and other extended emission sources known to be associated with Herbig-Haro objects are also discussed later.

It should be noted that several stars have saturated the detector during the 2 second on-chip integration, including BN, θ^2 Ori A, and several of the Trapezium OB stars. Unlike with optical CCD detectors, the DRO read out chip does not bloom or show column bleeding when saturated.

5.4.2 The H mosaic of Orion

In order to derive quantitative information about the energy distributions of the stars in either the visible Trapezium cluster, or any population of stars embedded in the background molecular cloud, extensive multicolour photometry is required for a substantial

fraction of the stars [65]. Therefore, in addition to the large K mosaic discussed so far, we also attempted to cover the same region through the broad band H filter, at 1.6 μm.

The H observations were made on the 18–19th of December 1986, with the other IRCAM instrumental parameters virtually identical to those for the K mosaic. Unfortunately, only half the mosaic was completed when an electronic problem injected pattern noise and many random bad pixels into the following images, making them unusable. It was subsequently discovered that all the previous data had also been affected by random bad pixels generated by the electronics. Secondly, after the data were reduced, it was found that severe optical ghosting made the frame to frame matching process very difficult. The filter causing this problem has since been replaced, but the H mosaic lacks the offset precision and carefully matched intensity values of the K mosaic as a consequence. Thirdly, the seeing was rather poor, with a FWHM of $\sim$ 2.0–2.4$''$, making it difficult to match the resolution of the K mosaic ($\sim$ 1.5$''$ FWHM).

However, the main problem with the H data is that the flat fielding technique was not used. After reducing the data in the same way discussed for the K mosaic, it was seen that the residual noise was a factor of four times higher than after applying the simple technique of raw source frame minus raw sky frame. As mentioned before, this simple technique does not flat field the data, and therefore is not likely to be accurate photometrically. Rather than abandon the H data entirely, we just present the final usable mosaic, shown in figure 5.19. Due the relatively poor nature of the data, no attempt was made to rotate, clean, and trim the mosaic. Further processing could perhaps improve the appearance of the image, but as further telescope time has been allocated to obtain full multicolour imaging photometry of Orion at J, H, K, and narrow band L, the effort involved is unjustifiable in light of the greatly improved camera performance since the data shown here were obtained.

5.4.2.1 General examination of the H mosaic

We can see that the limited region imaged at H is markedly different compared to the same region in the K mosaic. The dominant feature in the H image is the Trapezium cluster, clearly showing the bright OB stars at the core, with a spherical halo of fainter stars. In the K mosaic, the symmetry of this cluster is broken by BN and the bright

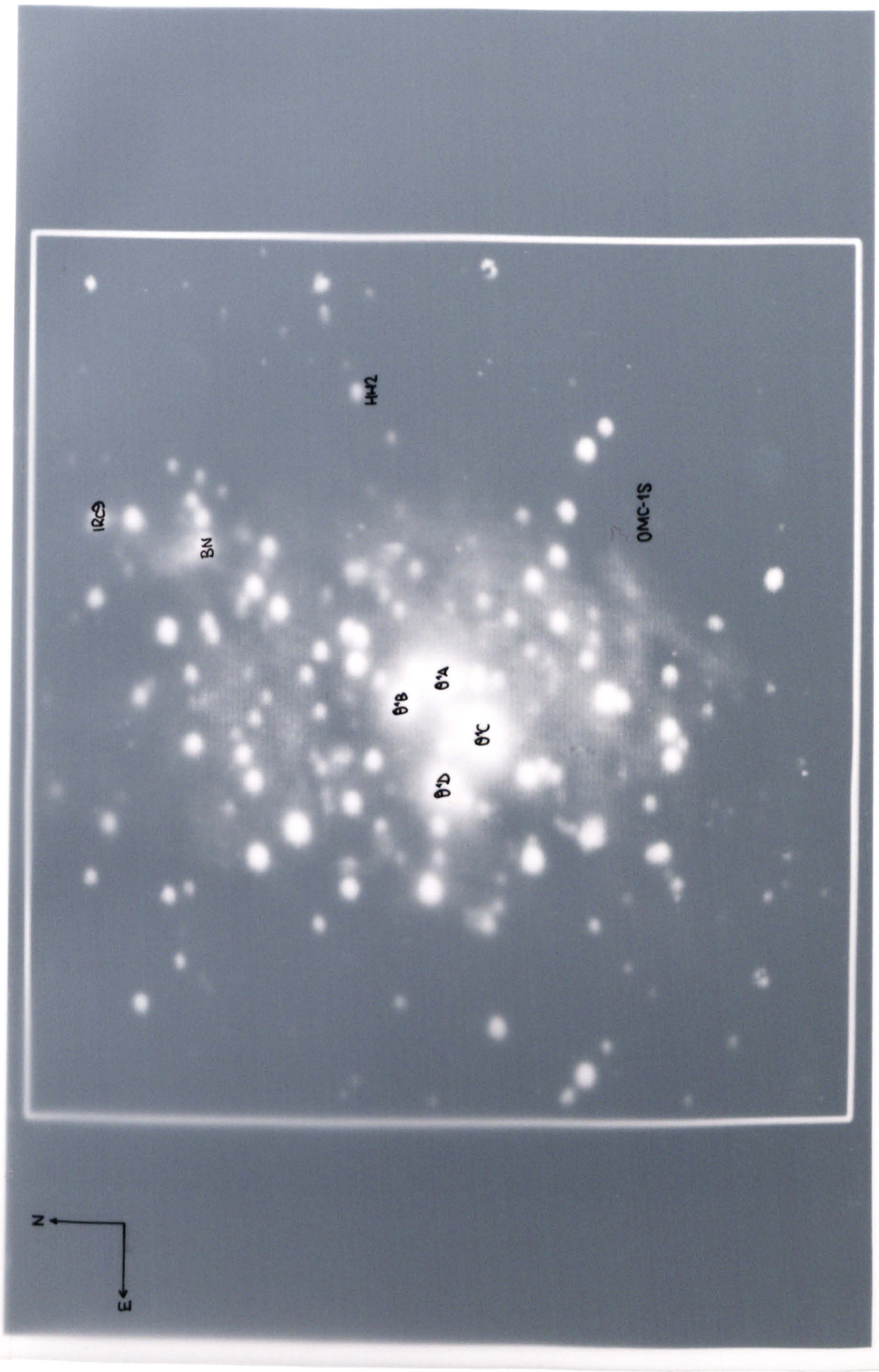


Figure 5.19: The Orion Nebula at H

N

E

IRC9

BN

HH2

θ'B

θ'D

θ'A

θ'C

OMC-1S

Figure 5.19: The Orion Nebula at H

reflection nebulosities associated with it and IRc2. BN is still visible at H, along with its reflection nebulosity to the north-east, although both are considerably fainter than at K. The optical stars around BN and IRc2 are all still clearly visible, but now appear to be part of the general Trapezium cluster, as opposed to being members of some secondary cluster around IRc2. Further discussions of this image are to be found below, when we look more closely at specific regions in the Orion Nebula.

5.4.3 Astrometry

Given the high spatial resolution and well defined pixel-to-pixel offsets that are available with arrays detectors, accurate astrometric positions are now possible in the near infrared. With this in mind, we have undertaken astrometry in the Orion K mosaic, with the following motivations :

- to re-define the local astrometric frame of reference, circumventing the previous dependence of authors on that defined by Parenago [117], thus accounting for any reference star proper motions that may have occurred over the past thirty years
- to obtain accurate positions for known and newly discovered point and extended sources, in preparation for follow-up imaging and spectroscopic work
- to assess the accuracy of the mosaicing techniques employed to generate a large scale infrared image from many individual frames
- to accurately define the actual pixel scale obtained when imaging in the (nominal) $0.6''$/pixel mode

5.4.3.1 Optical astrometry from Schmidt plates

The initial step taken was to link into a standard astrometric frame of reference. For this purpose, we required wide field coverage of the region surrounding the Orion Nebula, which not only showed a number of bright astrometric standards, but also a significant number of stars within the bounds of the IRCAM K mosaic which could serve as secondary standards. These two requirements are strongly conflicting, due to the intense brightness of the Orion Nebula itself. A very short exposure time would allow several

of the very brightest Trapezium OB cluster stars to be seen, without over-exposing the nebulosity. However, such a short exposure would not reveal any of the fainter stars required as secondary standards, nor many of the primary astrometric standards. Conversely, a longer exposure time would allow the primary standards to be clearly seen, whilst the Orion Nebula would be completely over-exposed.

The solution was to obtain Schmidt plates taken through narrow band interference filters and broad band plates taken in the photographic infrared ($\lambda \sim$ 8000–11000Å), which give both wide field coverage and a degree of nebular suppression, making it possible to see many primary standards in the far field *and* stars deep into the Orion Nebula on the same plate. A series of such plates was commissioned from the UK Schmidt Telescope (UKST) in Australia, with details as listed in table 5.3. Several filter, emulsion, and exposure time combinations were used by the UKST staff in order to meet the exacting requirements. In addition to the astrometric application, these same plates will also be used to widen the wavelength baseline of photometric measurements of the Orion Nebula region. Note that, as these plates were obtained at essentially the same epoch (2–4 months later) as the IRCAM data, proper motions between the photographic and IRCAM images can be entirely ignored.

For astrometric purposes, one plate (Plate WI11642) was found to show both standard stars in the far field and stars close to the Trapezium, within the bounds of the IRCAM image. Using the Starlink CHART package, the Perth Astrometric Catalogue was searched for all within a 3.5° box[1] centred on the Trapezium. The Perth Catalogue is the most accurate available, and is used to define the FK4 fundamental astrometric reference frame. Some 18 primary standard stars were unambiguously identified on the Schmidt plate. Several others were rejected due to confusion — the V band stellar magnitudes output from the catalogue are often poor discriminants when looking for the stars on near infrared photographs. At the same time, 31 additional stars close to the Trapezium on the Schmidt plate *and* visible on the IRCAM mosaic, were selected to serve as secondary standards.

Using the ROE Packman X–Y measuring machine, the positions of the 18 primary astrometric standards and 31 secondary standards were measured during one session.

[1]Only the central 3.5° of the 6.5° field of view of the Schmidt plate was exposed, due to the limited size of the photographic infrared filters

Plate number	Emulsion	Filter	Date	Exposure time (mins)	Notes
IF11641	IIIaJ	AAO640	18–19/2/87	9	Reasonable dynamic range
WI11642	IVN	RG830	18–19/2/87	15	Used for astrometry
IF11643	IIIaJ	AAO540	18–19/2/87	9	Bad seeing—little visible
Z11645	IZ	RG1000	21–22/2/87	120	Nothing visible at all
WI11647	IVN	RG830	22–23/2/87	3	Little visible in far field
WI11678	IVN	RG830	8–7/3/87	60	Reasonable—bad seeing
WI11710	IVN	RG830	12–13/3/87	60	Deeper than above—good

Table 5.3: Near infrared Schmidt plates of the Orion Nebula

The Starlink ASTROM utility was used to derive a six coefficient (x, y offsets; x, y scales; rotation; shear) plate solution for the Perth standards. This plate solution was then used to derive the positions of the nominated secondary standards. The RMS error between the actual and calculated positions for the Perth standards was only $\sim 0.83''$, very close to the absolute measurement accuracy limit of the Packman X–Y machine.

5.4.3.2 Astrometry in the IRCAM K mosaic

At this stage, we had fairly accurate ($\pm 0.8''$) astrometric positions for some 31 visible stars within the boundaries of the IRCAM K mosaic. Taking these as secondary standards, we proceeded in an analogous fashion to that outlined in the previous section. Here though, an ARGS screen and software replaced the Packman X–Y machine. The IRCAM image was displayed, and a cursor was used to roughly define the location of each secondary standard, after which a centroiding algorithm (CENTROID in Rapi2d) was used to more precisely locate the stars. The accuracy of the centroiding algorithm is approximately $\pm 0.2''$ in the $0.6''$/pixel mode, and centroiding errors contribute little to the overall error for this image.

Taking their astrometric positions as calculated from the Schmidt plate in the previous section, the ASTROM utility was used again, to generate a six coefficient 'plate' solution for the IRCAM image. At this stage, 5 of the 31 stars with large residual errors were rejected from the fitting procedure. The positions of three of these stars were poorly defined due to software centroiding problems — the first lies on a particularly intense patch of nebulosity; the second is one of a small cluster of stars; the third was simply not located by the centroiding algorithm. Of the remaining two, one is θ^1 Ori C, and was poorly measured on the IRCAM image due to saturation, and one was rejected due to a known slight 'slippage' in that region of the IRCAM mosaic.

After removing the five poorly fitted stars, the error in the calculated fit to the secondary standards was $\sim 0.6''$ RMS, with no single star more than $1''$ in error. The derived plate scales were $0.63''$/pixel in the x direction, and $0.62''$/pixel in the y direction, giving a mean plate scale of $0.625''$/pixel, well in accord with the nominal result used prior to obtaining these data. As mentioned earlier, these results were also used to determine that the mosaic was rotated $\sim 1°$ from north-south, east-west.

5.4.3.3 Astrometric accuracy

There are three astrometric figures of merit arising from this work :

- internal accuracy within one single 62 × 58 image
- overall accuracy of the whole 126 frame, 506 × 448 pixel mosaic astrometric reference frame
- accuracy of link into the FK4 fundamental astrometric reference frame

As IRCAM uses a physically highly stable detector array, the first figure is simply limited by the optical point spread function, and the software used to define image centroids. Isolated point sources imaged at $0.6''$/pixel in $1.5''$ seeing conditions can be located to within $\pm 0.2''$, and thus the relative positions of two point sources within the same image should be defined to $\pm 0.3''$.

The overall accuracy of the entire 126 frame mosaic is roughly defined by the mean error in the plate solution derived for the secondary standards measured in the mosaic. This error is $\sim 0.6''$, the size of one pixel. Given that only integer pixel offsets were used in the mosaicing process described earlier (section 5.3.2), this result is as good as is possible, and excellent justification of the intensive work that went into the mosaicing process.

Finally, the accuracy of the link into the FK4 astrometric reference frame is given by combining the errors resulting from the Packman and ARGS measuring processes. The error in the former was $\sim 0.8''$ and in the latter $\sim 0.6''$, giving us a total RMS error in the positioning of any given source in the IRCAM mosaic relative to the FK4 reference frame of $1''$.

As the IRCAM K image has higher spatial resolution and larger area coverage than any previous infrared mapping or imaging of the Orion Nebula, and given the high degree of astrometric accuracy, we feel it worth presenting both B1950.0 and J2000.0 equinox coordinates for a number of important sources in the mosaic. Where possible, point source positions were defined by the centroiding algorithm, otherwise the peak intensity pixel was used. The latter is also true for the selection of extended sources listed. The astrometric results are given in Appendix A.

5.4.4 Point source profile

The point spread in the K mosaic has a $\sim 1.5''$ FWHM, with a slight elongation in the east-west direction due to an oscillation in the UKIRT RA drive. Atmospheric seeing and telescope oscillation are probably the main components of the point spread profile, as the IRCAM optics have a spot size smaller than $1''$ in this mode, and seeing profiles of better than $0.8''$ FWHM have been demonstrated on the UKIRT with an optical CCD camera [10], indicating that the telescope is reasonably well figured. An additional factor to be considered is the mosaicing process, which has distorted the point source profile of many stars, but without significantly broadening the point spread profile in general.

The H data has a considerably worse seeing profile, with $2\text{–}2.5''$ FWHM typical across the image. Again, slight elongation due to the RA oscillation is seen in the east-west direction. It is known from measurements of bright standard stars through the H filter used at the time, that considerable ghosting occurred. Although it is not easy to see any such ghosts in a complex region such as Orion, at least three spurious sources have been identified in the H mosaic, in that they only appeared in one of the two mosaic frames in which they supposedly existed. In addition, although the view can change dramatically between H and K, as embedded stars are revealed at the longer wavelength, it would be very peculiar to see a given point source at H and not at K, assuming similar sensitivity limits.

5.4.5 Photometry

Due to the nature of the instrument commissioning process, both the K and H mosaic images were obtained without corresponding standard star data. Whilst this may appear to be an unfortunate oversight, the main intention at both times was to provide qualitative examples of the data that can be obtained with infrared cameras, rather than data for precise astrophysical experiments.

Nevertheless, due to the large number of stars covered by both mosaics, we are able to derive approximate photometric zero points for both data sets, by comparison with stars in common with the work of LBLS and HABW.

5.4.5.1 Photometric techniques

It is clear from both mosaics that many of the point sources are embedded in the bright and highly structured nebulosity, and therefore that stellar photometry will not be trivial, as this requires the accurate removal of any nebulosity, in addition to the sky background.

Two separate photometric techniques were used by LBLS and HABW. The data presented by the former were obtained using tangent beam chopping, which should suppress extended emissions. However, due to the complex nature of the limited region they surveyed, this technique will lead to errors where the emission is highly structured, or when two stars are close and confused in the 3.5$\hat{\pi}$ aperture used. HABW do not explicitly state how their point source photometry was performed, but they refer to 'net peak counts' as a measure of brightness before converting to magnitudes, in their Table 1. From this, the fact that their photometry agrees well with that of LBLS, and from discussions with one of the authors not involved in the data reduction, it is understood that some form of nebular background subtraction was done, by examining a strip chart of the raster map output, and by removing any reasonably slowly varying component from beneath the stellar profiles (Ward, private communication).

Neither technique is completely satisfactory, and with the improved spatial resolution, we should be able to do much better in principle. In a manner similar to that used by Herbig and Terndrup to remove the effects of nebulosity from their optical CCD photometry [65], complex point source profile fitting software was applied to the IRCAM data. However, the software package used (DAOPHOT) gave unsatisfactory results, mainly due to the variable point spread function across a mosaic. As each mosaic is comprised of many separate images, the seeing varies across it, and as the mosaicing process used at this stage only works in terms of integer pixel offsets, the point source profile of stars imaged more than once becomes randomly distorted.

Aperture photometry was also tried. For isolated stars, aperture photometry uses multiple annuli around the star to accurately estimate the sky flux level underneath it. However, due to the bright nebulosity and crowding, this procedure was not possible either. Finally, a somewhat crude form of aperture photometry was employed. The stellar flux was measured with a software aperture placed over the star, and the un-

derlying background flux was estimated by measuring the surface brightness of nearby nebulosity that was judged to be similar to that beneath the star. Repeating this process several times, a fairly reasonable estimate could be made of the stellar flux alone. This technique is relatively subjective, but was felt to be the best possible compromise at this stage, in this complex and highly structured region. Indeed, Terndrup has compared the results obtained with a similar technique to those obtained by profile fitting for their optical data, and found that although the errors were generally larger in the former case, there was no systematic shift in the derived stellar magnitudes (Terndrup, private communication).

Thus, we feel that the photometric data presented will not be systematically in error, but that the random errors may be quite poor. In general, after five repeated measurements of a star, an error of $\pm 0^{m}.2$ could be reached. However, it must be pointed out that this is in no way reflects the absolute photometric accuracy of IRCAM. Tests on isolated standard stars using more rigorous aperture photometry generally results in errors of $\pm 0^{m}.01$ for stars in the range $K = 7^{m}.0$–$14^{m}.0$, and the poor errors seen here are due to the special nature of the problem in the Orion Nebula.

5.4.5.2 Photometric calibration of the K and H mosaics

We have performed aperture photometry as described above, for many of the stars in common between our data and that of LBLS and HABW, in order to derive zero points for our data.

The results at H are simplest. Only LBLS have H data, and comparing 12 of their 26 sources with our data, we derived a zero point for the H mosaic with an error of $\pm 0^{m}.13$. Of the other 14 sources, several are unseen at H, several are actually extended emission knots, and several are confused with other nearby sources. At H, the BN-KL region is relatively free of extended emission, and we consider the LBLS magnitudes reasonably reliable.

At K, the situation is more complex. We have measured 11 of the LBLS sources, the others being too confused or extended to be included. From this we derive a zero point of $23^{m}.91 \pm 0^{m}.32$. Using the data of HABW, the same stars give the same zero point, as expected from the close correspondence between their data and that of LBLS.

However, measuring another 16 stars seen by HABW, outside the area covered by LBLS, we derive a zero point of $23^{m}.52 \pm 0^{m}.23$. Whilst it is possible that these two zero points just about agree within the errors, we feel that the latter zero point is more accurate for the following reason. The majority of the stars measured to derive it were isolated and were all outside the complex region surrounding BN-KL. Thus, our measurement errors were lower, and the effects of incorrect nebular subtraction lessened. Measurements within the BN-KL region are less likely to be correct. Indeed, considering the two very different techniques used by LBLS and HABW, and the complexity of the emission in the BN-KL region, we are surprised at the level of agreement between their photometric data. Finally, there is the possibility of error due to the fact that no atmospheric extinction corrections were made in the current work. However, the maximum airmass difference between any of the stars measured for photometric calibration purposes is less than 0.5, and with a nominal K extinction of $0^{m}.08$/airmass, this will not account for the observed discrepancy.

Thus, throughout the following sections, we shall refer to H and K magnitudes. The zero point for the H data was unambiguously obtained from the data of LBLS, whilst the zero point used at K was that determined from the more isolated stars measured by HABW, outside the BN-KL region.

5.4.5.3 Calibration for emission line nebulosity

In addition to point sources, considerable nebulosity is visible in the K mosaic. This is due to continuum emission and reflection nebulosity, and flux from various emission lines in the K window. In particular, we might expect to see the Brγ line of ionised hydrogen at 2.166 μm associated with the HII region, and lines of molecular hydrogen (e.g. the v=1–0 S(1) line at 2.122 μm), where there are shocks, for example.

Before we examine the nebulosity, we need to provide some rough estimate of the nebulosity flux density. In the absence of any precise calibration data, we can only derive such an estimate via an *ad hoc* method. Firstly, from the K point source photometry, we know the number of data numbers (DN) that is detected from a star of given magnitude in the K mosaic. This gives us a DN–magnitude conversion. Next, we use SIRCAM to predict the number of photo-electrons per second that should be detected from a $0^{m}.0$

star at K, and also from a line source of brightness $10^{-20}\,W/cm^2/\square''^2$ at $2\,\mu m$, in the relevant IRCAM observing mode. Combining the two resulting numbers with the DN–magnitude conversion, we have derived an approximate DN–flux density calibration for the K mosaic. As a cross check of this crude calibration, we can use the figure quoted by Hayashi *et al.* [62], that a $K = 6\overset{m}{.}45$ star corresponds to a flux density of $10^{-16}\,W/cm^2/\mu m$ at $2.2\,\mu m$. Assuming a width for the K filter of $0.5\,\mu m$, this figure agrees with our crude calibration for the K mosaic to within 20%.

5.4.5.4 Sensitivity limits

From the photometric calibration of the mosaics, we are now able to calculate the limiting sensitivity to both point sources and extended emission in both data sets.

The very faintest point source visible in the K mosaic has an upper limit magnitude of $K = 15\overset{m}{.}1$, and is located in the region to the south-east of the bright bar, where the background is smooth and free of nebulosity. The star is just visible above the sky background noise. Over most of the mosaic however, the sensitivity limit to point sources is not set by the sky background noise, but by variations in the extended emission masking faint point sources beneath. This is particularly true towards BN-KL. IRc2 is just visible as a point source in the K mosaic, and in its one measurable pixel above the noise, IRc2 has a brightness of roughly $K = 15\overset{m}{.}0$.

Measuring the RMS noise in the south-east corner of the image, we derive a limiting 1σ surface brightness sensitivity of $K = 18\overset{m}{.}1$ per pixel, or $17\overset{m}{.}1$ per square arcsecond. Again however, the limit across most of the image is set by the variations in the extended emission, rather than the sky noise. In terms of an emission line brightness, the corresponding 1σ flux density limit for the image corresponds to $5 \times 10^{-21}\,W/cm^2/\square''^2$ at $2\,\mu m$.

Both the point source and extended emission sensitivities fall well short of the predicted sensitivities for this mode and integration time. From the results presented in chapter 4, we might expect to obtain a 3σ point source detection in the $0.6\,\square''$/pixel mode in 1 minute of about $K = 17\overset{m}{.}0$, and a 1σ limiting surface brightness of about $K = 19\overset{m}{.}2/\square''^2$. The reason for this apparent shortfall in sensitivity is again due to the excess dark current and photon leakage in these data, and we would hope to achieve sub-

stantially better results with the camera when this experiment is repeated in December 1987.

As we have already mentioned, the quality of the H mosaic is poor, and the faintest point source visible in it has a magnitude of approximately $H = 13^{m}.2$, much worse than predicted. It is worth noting that the seeing in the H mosaic was about 2–2.5″ FWHM, and this image blurring lowers the point source sensitivity considerably. Due to the large numbers of bad pixels towards the edge of the image, we were unable to estimate a limiting surface brightness.

5.4.6 The K frequency function

The density of stars in the Trapezium cluster is very high, as noted by many optical observers, and as summarised by Herbig [64]. However, studies of this young dense cluster are considerably hampered in the optical due to the intense and highly structured nebulosity associated with the HII region. Only recently have studies of the photometric properties of the cluster been possible, through the use of CCD cameras, which have high enough quantum efficiency to enable studies through narrow band filters, blocking out the emission lines of the nebulosity. In addition, the linearity of CCDs and their high dynamic range enable the use of sophisticated software techniques to remove residual contamination due to the highly structured background. Herbig and Terndrup [65] used these techniques to obtain accurate photometry of almost 100 stars in the central region of the cluster, all of them measured through a narrow band filter near 8000Å and two thirds also measured near 5500Å. After drawing conclusions on the nature of the cluster, Herbig and Terndrup stressed the need for multicolour infrared photometry for as many members as possible, in order to assess the energy distribution of the stars, thus deriving the luminosity. Also, infrared measurements penetrate further back into the molecular cloud. This will disclose whether or not the optical cluster is part of a pre-existing distribution of stars pervading the cloud, only recently revealed in the cavity excavated by the Trapezium OB stars on the front face of the cloud.

With these points in mind, and given the extensive spatial coverage of our K mosaic, we have derived a K frequency function for the region. Over 440 stars were measured using the software aperture technique discussed above, and calibrated using the zero

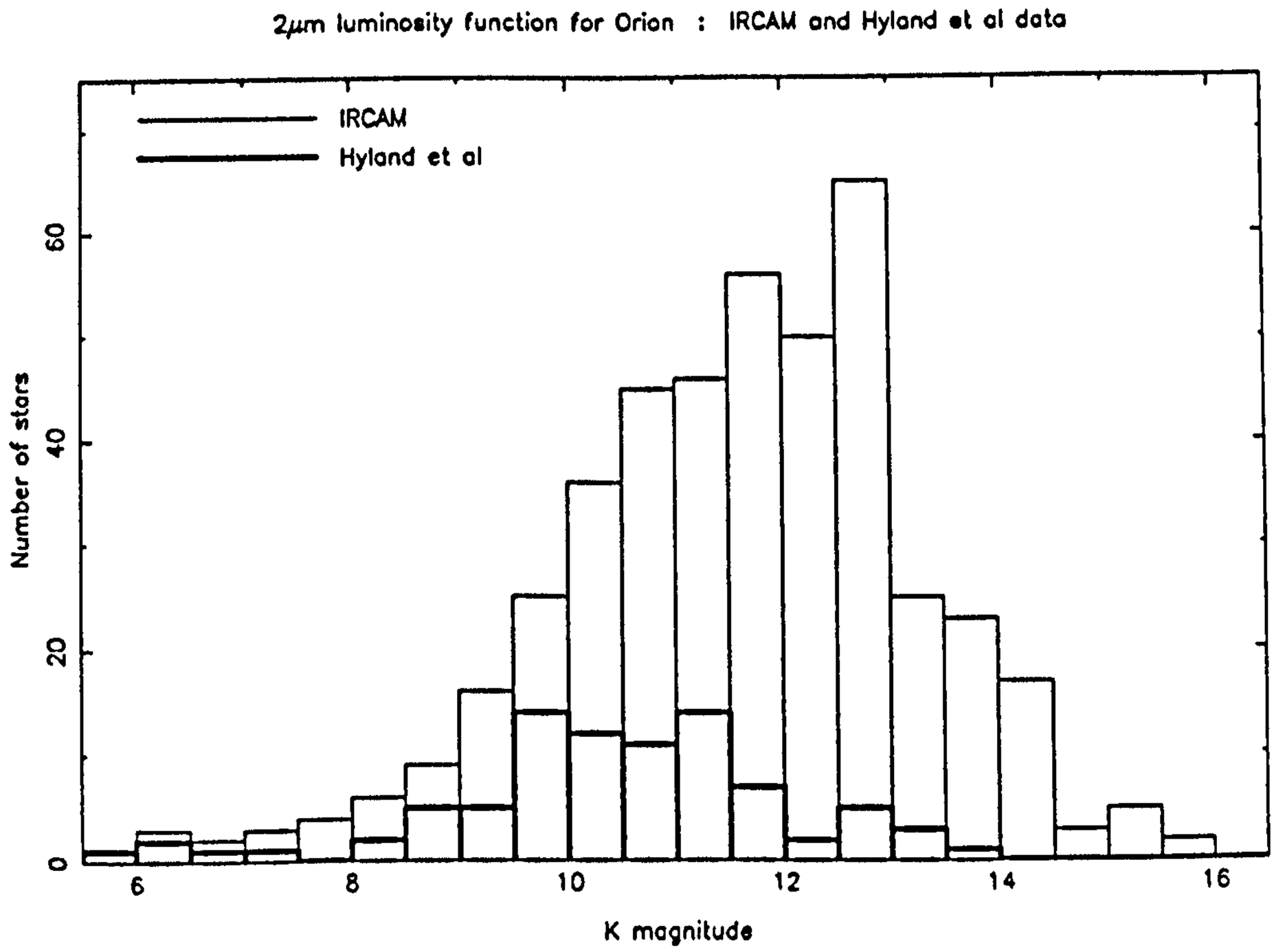


Figure 5.20: K frequency function for the Orion Nebula

point derived from the data of HABW. As discussed earlier, we attempted to use the more sophisticated software techniques of Herbig and Terndrup to provide more accurate photometry, but were unable to obtain consistent results, mainly due to problems with image mosaicing. However, as the errors in the magnitude measurements are essentially random, their cumulative effect should not systematically affect the frequency function. Thus, whilst we feel it unwise to present the photometry of each individual star,given the large errors, the frequency function should be an accurate reflection of the larger scale distribution.

The derived frequency function is given in figure 5.20, along with the (rebinned) frequency function determined by HABW. Their frequency function was derived from 88 stars, compared to our 440. As a consequence of the larger number of stars, the IRCAM frequency function is far smoother than that of HABW. Although the numbers of stars measured will obviously affect the absolute height of the histogram bins, the basic shape should be similar, if the frequency function is roughly constant spatially. The two frequency functions clearly do not have the same shape, with the IRCAM data having a pronounced cut-off about a one and a half magnitudes fainter than that of HABW, and a larger fraction of fainter stars overall.

As stars in common between the two data sets were used to derive a fairly accurate zero point for the IRCAM mosaic, the difference does not lie in incorrect photometric calibration. For the same reason, neither does it lie in the use of different photometric techniques. As stated before, inaccuracies in the IRCAM magnitudes should be random, and thus should not affect the shape of the frequency function. At their respective cut-off magnitudes, both data sets are almost two magnitudes above their sensitivity limits, and thus we are not seeing a sensitivity effect. Finally, the lack of atmospheric extinction corrections would account for a faintward shift in the IRCAM frequency function of only $0^m.1$–$0^m.2$ at the most.

We can only assume that the frequency function is *not* spatially uniform, and that we have probed a different cross section of the total population of stars than HABW. They measured 88 stars in a region of $1' \times 4'$ centred on BN, the upper section of which is not covered by our image. Our image covers a region $5.3' \times 4.7'$ centred on θ^1 Ori C, and in this region we detect some 440 stars. Thus, HABW see a stellar density of approximately 22 stars per square arcminute, whilst IRCAM sees a somewhat lower density of about 18 stars per square arcminute. This is simply accounted for by the fact that the IRCAM mosaic extends much further outside the core region observed by HABW, and that we are seeing out towards the edge of the cluster. By examining the K mosaic alongside the region covered by HABW, it is obvious that the frequency functions are spatially different. Particularly to the east of BN, in the dark bay region, and to the east of the Trapezium, many faint stars are seen in the IRCAM data, whereas the region covered by HABW, to the south and north of BN, appears to contain a larger proportion of brighter stars.

At this stage however, on the evidence of the K image alone, we cannot decide between several possible reasons for this spatial difference in frequency function. It may be that with the extended spatial coverage of the IRCAM data, we are seeing a halo of lower mass and luminosity stars, that have formed outside the central core region of the Trapezium cluster (Zinnecker, private communication). Alternatively, we may simply be measuring the effects of localised foreground extinction in the dark bay region, on otherwise normal members of the Trapezium cluster. As HABW did not measure stars in the dark bay, their frequency function would naturally cut-off at brighter magnitudes. A third possibility is that the Trapezium cluster extends back into the molecular cloud. In the IRCAM data we would be seeing the cluster extending back into the less opaque

regions of the background extinction, to the east and west of BN-KL. As HABW only examined the region along the molecular ridge to the north and south of BN-KL, they might not see these additional stars, as the heavier extinction nearer the core of the ridge could effectively block the light from them completely.

Herbig and Terndrup used their optical CCD photometry along with the frequency function of HABW to examine the possibility of the Trapezium cluster extending back into the molecular cloud. Their procedure was as follows (Herbig, private communication). Excluding the Trapezium OB stars, all stars in their optical data which had also been measured by HABW at K were drawn up into a histogram according to their K magnitudes. These were assumed to be the stars in the front slab of the molecular cloud only. Then, for a series of 0.44 pc slabs, each with certain number of magnitudes of K extinction, the foreground histogram was reddened accordingly, and added to a cumulative frequency function. This was repeated until the whole depth of the cloud had been included. The final frequency function was then compared with the actual K frequency function as measured by HABW.

Two parameters are to be chosen here; the total depth of the cloud, and the extinction per slab. Herbig and Terndrup choose a total depth D of 4.0 pc and an A_K per slab of $4\overset{m}{.}0$ as being reasonably consistent with molecular and other infrared data. Using these plausible parameters, Herbig and Terndrup have derived a frequency function that predicts 3–5 times as many stars fainter than K = $9\overset{m}{.}5$ than actually seen by HABW. Only in the limits of a very dense cloud, where no stars are seen at K beyond the first slab, or a very thin cloud, where there is very little additional cloud behind the optical cluster, do the predictions begin to approach the actual K frequency function of HABW. From this, Herbig and Terndrup conclude that the optical cluster is confined to the front face of the molecular cloud only, and does not extend back into it.

Due to the large number of stars covered in our K image, the relatively inaccurate infrared photometry obtained so far, and the inadequate optical photometry to cover this region, we are not able to perform a similar comparison between our observed K frequency function, and a predicted frequency function based on optical data. We hope to do so after reobserving the region with enhanced sensitivity, and at J, H, K, and nbL this coming winter. At this point, we are only able to note that there *are* a considerably larger number of faint stars seen in the IRCAM frequency function than in that of

HABW, and that the question of the spatial distribution of the cluster members, both in the line of sight and across the face of the region, must remain open at this point.

5.4.7 Comparison with Lonsdale et al. and Hyland et al.

When HABW applied the superior technique of DC raster mapping to the inner region of the Orion Nebula, the improvements in apparent resolution and sensitivity to extended emission were such that they were able to re-examine the tangent beam AC mapping by LBLS of the same region, even though both maps were made with the same circular aperture of $3.5''$ diameter. Now that we have imaged this region with IRCAM with $0.6''$ pixels, and seeing limited resolution of about $1.5''$ FWHM, a further re-evaluation of the morphology of both previous maps is now in order.

We shall concentrate on the $40'' \times 40''$ region centred on BN, mapped by LBLS at H, K, and L, and by HABW at K only. It is worth noting that this region is approximately that contained within one IRCAM image, and thus that the relative position accuracies within this image will be very high.

LBLS discussed 26 sources, labelled alphabetically. Twenty three of these sources were described as being point-like, with profiles no wider than that of a standard star. Two of the remaining sources (f, j) were confused with brighter nearby sources, and source q was noted as being definitely extended. Eleven sources were identified with visible stars, leaving the remaining 15 as infrared identifications only. As 26 sources were detected in only 120 resolution elements, LBLS described their search as confusion limited, with the caveats that (i) some sources may be multiple and (ii) that local enhancements in extended emission may be mistaken for point sources due to the small chop throw used.

HABW confirmed the LBLS sources $a, e, g, h, k, t, v, y, z$ as point-like and associated with visible stars. Sources b, c, u are only suspected in their map, and thought to be visible stars, whilst j is also suspected, but with no visible counterpart. The 'pure' infrared sources d, i, and BN are confirmed. However, at K, HABW were unable to detect sources r and s, the former seen only at L by LBLS, and the latter only at H. Sources m and p were not seen either, despite their detection at K by LBLS, whilst HABW considered the sources f, l, n, q, w, x to be enhancements in the extended

emission only.

With the high spatial resolution of IRCAM, the following identifications can be made. Sources $a, b, c, e, g, h, k, t, u, v, w$ and z (double) are all clearly seen to be point-like, and are associated with visible stars on the AAT I plate. The infrared source BN dominates this region at $2\,\mu$m, and also clearly visible in the IRCAM image are m, p, s, and r (just), the latter four unseen by HABW. Sources i and d are not point-like, and source j is extended structure, contrary to the point source identifications of these three sources by both LBLS and HABW. In agreement with HABW, sources f, l, q, w are seen to be only extended emission, but contrary to HABW, sources n and x *are* seen to be point-like at this high resolution.

Thus, IRCAM detects all 26 of the LBLS sources : twelve are point sources associated with visible stars ($a, b, c, e, g, h, k, t, u, v, y, z$); seven are pure infrared point sources (BN,n, m, p, r, s, x); and the remaining seven are either nebulous sources (i), or just extended emission enhancements (d, f, l, j, q, w). These results are summarised in table 5.4, and the various sources are identified in figure 5.21.

One important point immediately arises from this comparison. On the strength of their analysis, HABW felt unable to confirm the claim of LBLS that a cluster of low mass, embedded, young stars exists in the vicinity of the higher mass sources, BN and IRc2. HABW contradicted this claim on the grounds that, of the twelve 'pure infrared' point sources (including BN and IRc2) discussed by LBLS, they had been able to confirm only three (including BN) as visible at $2\,\mu$m and point like. However, the new IRCAM data indicates that at least seven of the LBLS sources *are* visible and point like at $2\,\mu$m. Excluding BN and IRc2, we are able to confirm five of the ten new sources found by LBLS, as opposed to the two confirmed by HABW. Therefore, the low mass, embedded, young star cluster hypothesis should probably remain open until further high resolution multicolour imaging and spectroscopic data can be obtained for these sources.

In the H mosaic, several of the LBLS infrared sources are visible, in addition to BN. Sources m, p, s, and x are still seen, although r (IRc2), only very faintly visible at K, is definitely not seen at H. Interestingly, much of the reflection nebulosity, seen at K to the south-west of IRc2, is not seen at H, and in fact source k, identified as an optical star, but distinctly nebulous at K, is clearly seen at H as a point source. Source i, associated with IRc3, and again nebulous at K, is also more point like at H.

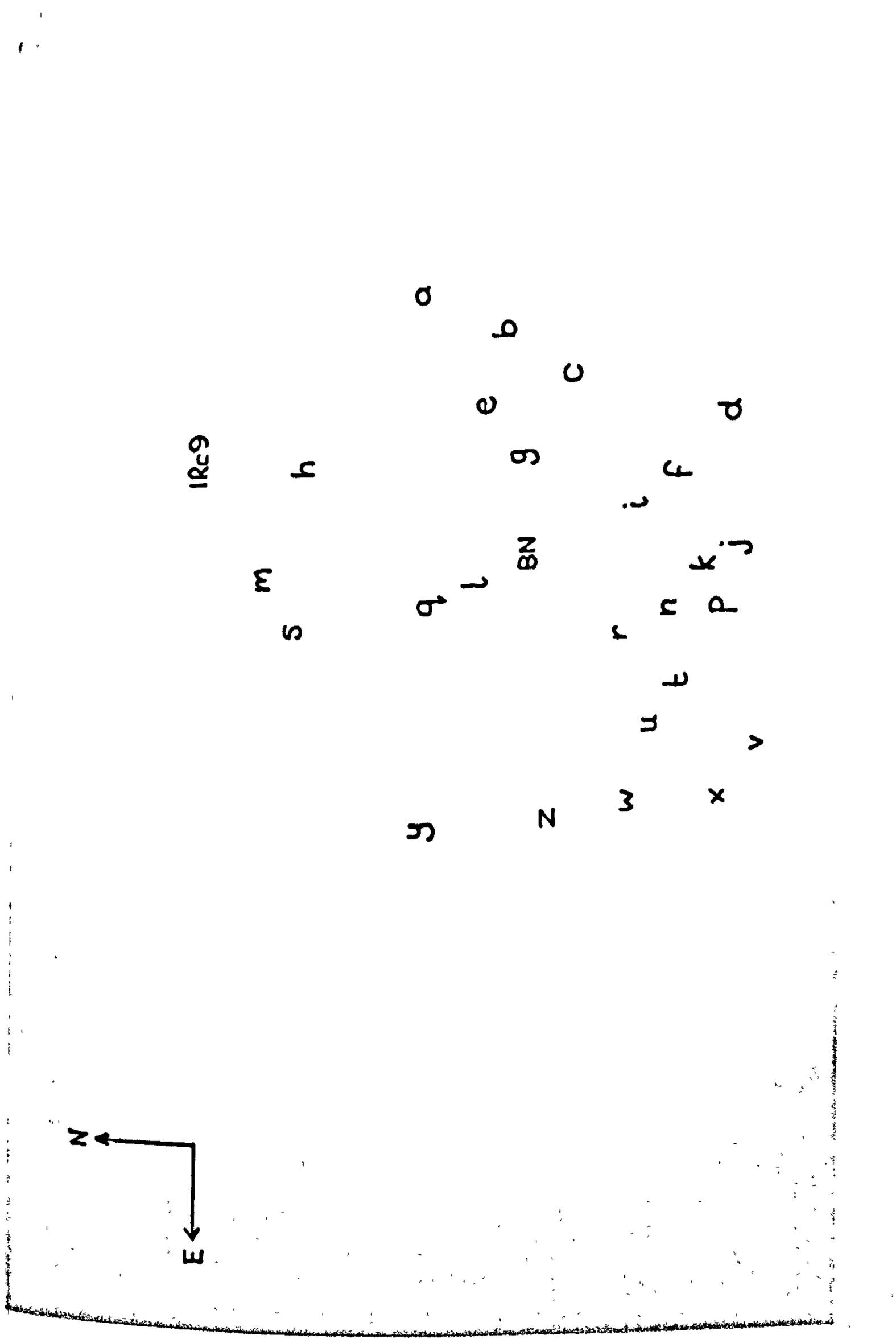


Figure 5.21: BN-KL and identifications of the Lonsdale *et al.* sources

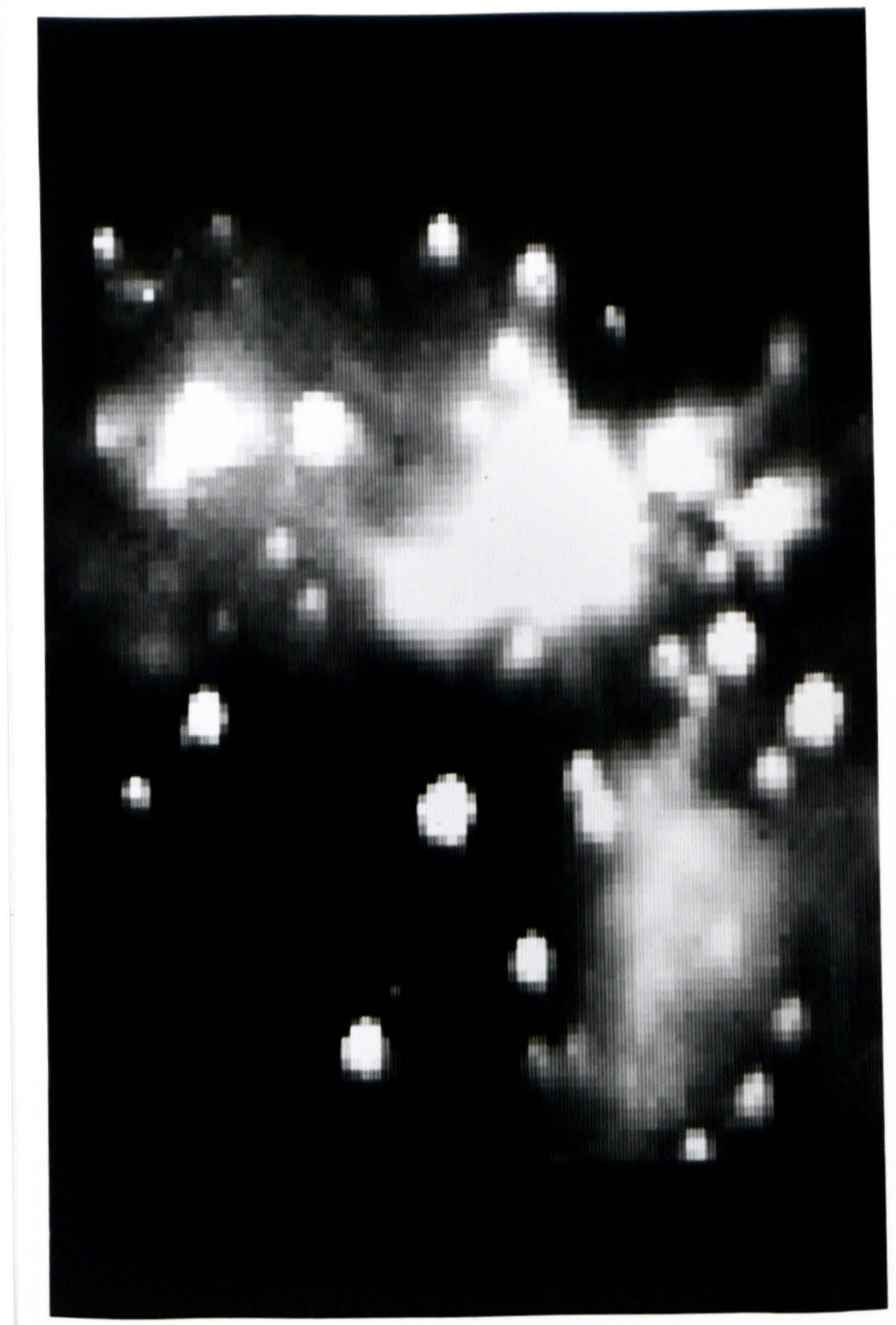


Figure 5.21: BN-KL and identifications of the Lonsdale *et al.* sources

Figure 5.21: BN-KL and identifications of the Lonsdale *et al.* sources

LBLS	Description	HABW	IRCAM	Other names
a	visible ⋆ ?	seen—⋆	seen—⋆	
b	visible ⋆ ?	suspect—⋆	seen—⋆	
c	IR only	suspect—⋆	seen—⋆	
d	IR only	seen—IR	structure	
e	visible ⋆	seen—⋆	seen—⋆	π1820
f	confused	structure	structure	
g	visible ⋆	seen—⋆	seen—⋆	π1821
h	visible ⋆	seen—⋆	seen—⋆	π1819
i	IR only	seen—IR	nebulous	IRc3
j	confused	suspect	structure	
k	visible ⋆	seen—⋆	seen—⋆	π1822
BN	IR only	seen—IR	seen—IR	IRc1
l	IR only	structure	structure	
m	IR only	not seen	seen—IR	
n	IR only	structure	seen—IR	IRc4(?)
p	IR only	not seen	seen—IR	
q	extended	structure	structure	
r	IR only	not seen	seen—IR	IRc2
s	IR only	not seen	seen—IR	
t	visible ⋆	seen—⋆	seen—⋆	π1839
u	visible ⋆	suspect—⋆	seen—⋆	
v	visible ⋆	seen—⋆	seen—⋆	π1840
w	IR only	structure	structure	
x	IR only	structure	seen—IR	
y	visible ⋆	seen—⋆	seen—⋆	π1838
z	visible ⋆⋆	seen—⋆	seen—⋆⋆	

Table 5.4: Comparison of LBLS, HABW, and IRCAM data

5.4.8 The southern molecular peak — OMC-1S

In this section, we shall examine a group of infrared point sources that are located very close along the line of sight towards a secondary condensation in the OMC-1 molecular cloud, some 90″ to the south of the well known peak in molecular emission associated with the deeply embedded source IRc2, and the BN-KL cluster of infrared sources, reflection nebulosity, and high velocity molecular outflow.

In contrast, it has been assumed for several years that the southern condensation, which is cooler and less luminous, but just as massive as its northern counterpart, has no associated point sources, the source of its luminosity being unknown. With our new infrared array images, recent tentative identifications of several embedded infrared sources in this region are confirmed, and ambiguity over which sources lie closest along the line of sight to the core of the molecular condensation is resolved using observations at H, K, and narrow band L. It is suggested that these new sources are embedded in, and responsible for the luminosity of the southern condensation, which we shall henceforth refer to as OMC-1S.

5.4.8.1 A general review

In the optical, the emission nebulosity associated with the foreground HII region dominates the Orion Nebula, whilst the stars of the Trapezium cluster and nearby BN-KL region dominate near infrared views, shortwards of 3 μm. Between 3 and 30 μm, extended thermal emission in the cores of both these clusters dominates. Beyond 30 μm however, extended emission from the younger embedded BN-KL region outshines the Trapezium. At far infrared, submillimetre, and millimetre continuum and line wavelengths, emission from the underlying molecular cloud OMC-1 is seen, and the fact that the peak emission at virtually all such wavelengths lies at the location of the BN-KL cluster is part of the overwhelming evidence that the BN-KL cluster lies within the molecular cloud, and that it is the BN-KL cluster, in particular IRc2, that is responsible for heating of the central core of OMC-1 to about 80 K [52, Table 3.1.III].

However, at high enough spatial resolution ($\sim 30''$), OMC-1 is seen to be structured, with several components lying along a ridge connecting OMC-1 to OMC-2, some 13′ to the north [16]. In addition to the structure extending northwards, the molecular cloud

exhibits considerable structure to the south of BN-KL also. Continuum observations of molecular clouds at far infrared and submillimetre wavelengths trace the density distribution of warm ($\sim$ 40–100 K) dust associated with the molecular gas. At an effective wavelength of 400 μm, and with a spatial resolution of 35″, Keene *et al.* [75] mapped the central core of OMC-1, revealing a previously unknown mass condensation some 1.5′ to the south of the condensation centred on BN-KL. This southern condensation, OMC-1S, was seen to be just as massive ($\sim$ 50–100 $M_\odot$) as the northern condensation, OMC-1N, but only half as luminous at 400 μm, with a correspondingly lower effective temperature of $\sim$ 45 K.

Subsequent high spatial resolution far infrared, submillimetre, and millimetre observations have shown a similar structure, which we shall now summarise.

A far infrared (45–160 μm bandpass) map of OMC-1 revealed OMC-1S after spatial deconvolution [36], and in combination with the 400 μm result, a total luminosity of $\sim 8 \times 10^3$ $L_\odot$ was derived for OMC-1S, in contrast to $\sim 8 \times 10^4$ $L_\odot$ for OMC-1N.

The J = 1–0 transition of ^{13}CO at 110 GHz (2.7 mm) remains optically thin at high densities, and a map of the ^{13}CO brightness can be used as a tracer of the molecular gas density, and thus the associated dust. OMC-1S is seen in the ^{13}CO line, but not in J = 1–0 ^{12}CO line (115 GHz, 2.6 mm), which is easily thermalised and optically thick at low densities ($\sim 10^3$ cm^{-3}), therefore tracing the gas kinetic temperature rather than density [125]. OMC-1N is seen in the ^{12}CO line, and not in ^{13}CO, confirming OMC-1N as a warmer source, and OMC-1S as a dense, cool condensation.

Formaldehyde (H_2CO) lines at 6 cm (1_{10}–1_{11}) and 2 cm (2_{11}–2_{12}) are most frequently seen in absorption, even against the 2.7 K microwave background, due to very low excitation energies and collisional pumping. However, at sufficiently high densities, the collisional pump is quenched, and the lines may be seen in emission. This is the case for both the 6 and 2 cm lines towards OMC-1N, inferring a high H_2 density there [72,15]. OMC-1S is seen in absorption at 6 cm [72], but in emission at 2 cm [15]. This discrepancy between OMC-1N and OMC-1S is explained by the difference in temperature between the two condensations [15], and one interpretation is that OMC-1S may be *inside* the Trapezium HII region, rather than behind or in front of it [160]. Briefly, as no optical absorption is seen towards OMC-1S, it cannot be in front of the HII region. If it is within the HII region, then ionised and optically thin material at a continuum temperature of

20 K at 6 cm will be at a temperature of 2 K at 2 cm — as the excitation temperature of the H_2CO K doublet is $\sim$ 10 K, this explains seeing the 6 cm line in absorption, and the 2 cm line in emission.

Ammonia (NH_3) metastable inversion lines require relatively high H_2 densities ($\geq 10^{4.5}\,cm^{-3}$) for thermalisation, in contrast to the lower densities required to thermalise the ^{12}CO J = 1–0 line, as mentioned above. Thus, these lines of NH_3 may be used in much the same way as the ^{13}CO J = 1–0 line, to examine the extent of higher density regions. Both OMC-1N and OMC-1S are seen in the (1,1), (2,2), and (7,7) inversion lines of NH_3 [16,93].

Continuum emission at 3 mm arises either from the free-free process, or from dust — in the M42 HII region, the former process is responsible for emission observed near the strong ionisation fronts, whilst the latter process dominates in the dense molecular clumps. Similar again to the ^{13}CO J = 1–0 line, continuum emission from dust at 3 mm is optically thin, providing another tracer of column density enhancements. At 3 mm, OMC-1S is revealed as one of several local massive ($\sim 50\,M_\odot$) condensations, all considerably in excess of their Jean s mass, and likely to undergo massive star formation in the very near term ($\sim 2 \times 10^4$ years) [114]. Observations in the CS J = 2–1 line near 3 mm, also sample high density material ($\geq 10^4\,cm^{-3}$), and high spatial resolution mapping in this line confirms the picture of a 'necklace' of massive condensations strung out along a line joining OMC-1N and OMC-1S [114,60].

Finally, emission from the SiO J = 2–1 line is thought to arise in the high velocity gas seen towards outflow sources and masers associated with high mass star formation [35]. Recent observations in this line have revealed that, in addition to the previously known emission peaking in the IRc2 outflow, a secondary emission peak is seen to the south, at the location of OMC-1S [169]. The high velocity wings (12–15 kms^{-1}) seen in the SiO line towards OMC-1S seem to indicate a very young, deeply embedded source.

To summarise : All the long wavelength (far infrared to radio) evidence reveals a compact ($\leq 30'' \times 30''$) condensation of some 60 $M_\odot$, as massive as OMC-1N, and in excess of its Jean s mass; a temperature of $\sim$ 50 K, and with a total luminosity an order of magnitude lower than that of OMC-1N ($\sim 8 \times 10^3\,L_\odot$). This points towards a likely site of future star formation, in much the same way as the recent formation of the KL cluster in OMC-1N, whilst the recent discovery of an SiO outflow in the region would

seem to indicate that star formation has already commenced.

5.4.8.2 The source of luminosity in OMC-1S

The vital question avoided thus far is : What is the source of the luminosity in OMC-1S ? As reviewed above, OMC-1S is thought to be a site of future star formation, and a natural candidate to explain the heating of the molecular condensation would be that star formation has already proceeded, and that one or several young, deeply embedded stars lie at the core of OMC-1S, as suggested by the recent discovery of a nearby SiO outflow. However, it has been generally accepted that, in contrast to IRc2 and the KL cluster at the heart of of OMC-1N, no such analogous cluster yet exists in OMC-1S.

Firstly, no stars are visible towards OMC-1S at optical wavelengths, but this same argument may be applied equally well to IRc2 and the KL cluster. Secondly, on discovering OMC-1S at 400 μm, Keene *et al.* mapped a $30'' \times 30''$ region centred there at 20 μm. No point sources were discovered to a limiting flux density of 60 Jy, with this limit set by confusion due to extended sources in the reference beam [75]. Thirdly, no point sources or extended emission enhancements had been observed towards OMC-1S in a $7.5''$ aperture at 2.2 μm to a limiting flux density of 13 mJy/beam, in a large scale raster map of OMC-1 [18].

Thus, in discussing the evolutionary status of OMC-1S, it has been generally assumed that no stars are present there, as summarised by Werner : "This condensation therefore appears to be a separate cloud fragment that has not yet begun to form stars, or within which the star formation process has been inhibited or terminated." [154]. This apparent lack of central luminosity source has been little discussed in the literature, but at least one model has had to invoke heating by the Trapezium stars to explain the observed luminosity [36] — however, the geometry required for the luminosity to observed at the longer wavelengths only make this suggestion somewhat unlikely. Clearly then, the source of the luminosity in OMC-1S is of critical importance in deciding its evolutionary status.

It is now clear that there are infrared point sources located close to OMC-1S. Several years before the far infrared discovery of OMC-1S, the Wyoming group made low resolution maps of the central region of OMC-1 at several wavelengths in the 8–13 μm

window [48]. It is worth noting (with the benefit of hindsight) that in their 10 and 12 μm maps, enhancements are indeed visible towards OMC-1S, although at such a low level ($\sim$ 6.5 mJy/$\square''^2$ $\sim$ 1/100th BN-KL) that no comment was made by the authors at the time. This indicates that the reason that Keene *et al.* did not see any point source emission at 20 μm was a merely lack of sensitivity.

Three near infrared point sources near the location of OMC-1S were first discovered by HABW in their high spatial resolution raster map at 2 μm, and briefly noted with regards possible recent star formation in OMC-1S. However, on combining substantially the same data with another raster map made at 3.6 μm, ABH noted the presence of a single very red point source some 15$''$ to the north of the cluster of three 2 μm sources, again briefly mentioning it in the context of a possible embedded source in the core of OMC-1S. Unfortunately, this latter colour composite image did not extend far enough southwards to include the cluster of three sources, and thus no indication of their colour was given. These colour composite images were presented in Sky and Telescope, and to date, no scientific papers have appeared giving details of these measurements, important in deciding which, if any, of these near infrared sources are associated with OMC-1S.

All these sources are clearly seen in the IRCAM K mosaic, and the cluster of three sources is also faintly seen at H. We have also managed to obtain a narrow band L image of the region, as part of the UKIRT service observing programme. We shall briefly discuss how the nbL data were obtained and calibrated before proceeding with our discussion.

5.4.9 Narrow band L imaging of OMC-1S

Under the UKIRT service observing programme, a proposal was made to obtain detailed follow-up imaging of the region the OMC-1S infrared sources at narrow band L (nbL, $\lambda = 3.6\,\mu$m, $\delta\lambda = 1.8\%$). Due to bad weather and the limited observing time available on Orion in March, only a very small amount of poor quality data was obtained. Nevertheless, it is worth briefly examining these data, as they extend the wavelength baseline further into the thermal infrared.

One image of the OMC-1S region was obtained at nbL, comprising 1000 coadds, each

with an on-chip integration time of 130 milliseconds, giving a total on-source integration time of only 2 minutes. A sky frame of similar length was also obtained, looking at the same 'blank' region of sky used whilst making the K mosaic. These data were reduced in the usual manner, flat fielding with the sky frame, although again, no corrections for the slight detector non-linearity were made.

For calibration, a standard star (HD129653) also observed at nbL at a similar airmass, was reduced in a similar way. Software aperture photometry was used to obtain an integrated stellar flux. Care was taken to normalise this flux correctly, as the on-chip integration time and number of co-adds used for the standard star were different to those used for OMC-1S. This process assumed the camera system is linear with respect to the number of co-adds, and with respect to the on-chip integration time. The former assumption is true, as is the latter in the heavily photon background dominated regime seen here. HD129653 has an L magnitude of $6^{m}.90$ and an L′ magnitude of $6^{m}.92$. As the star has a small L – L′ of $0^{m}.02$, we can safely assume an nbL magnitude of $6^{m}.9$, and this gives us a zero point for the OMC-1S image with an error of $\pm\ 0^{m}.03$.

5.4.9.1 OMC-1S at H, K, and nbL

In figure 5.22 we present single image views (62 × 58 pixels) of OMC-1S at H, K, and narrow band L, with an accompanying key to the sources. Several Parenago [117] stars are shown for reference. We have numbered the cluster of three infrared sources 1 to 3 in order of their K brightness, and the single very red point source to the north, between π1807 and π1808 has been numbered 4. Accurate positions for these sources are to be found in Appendix A, and we shall refer to them as S1–S4.

It is clear that all the sources are quite heavily reddened compared to the known optical stars. Sources S1–S3 are just visible at H, significantly brighter at K, and source S1 is also bright at nbL. We can see that source S3, the faintest source, just to the northwest of source S1, appears bluer than sources S1 and S2, as confirmed in the photometric results given below. There is also a suggestion that sources S1–S3 may not be quite point-like compared to the known optical stars, and that they may be embedded in some slight local nebulosity, as opposed to the general bright nebular background peaking to the east. Confirming the result of ABH, source S4 is very red indeed, being invisible at

Figure 5.22: IRCAM images of OMC-1S

Source	Magnitude		
	H	K	nbL
S1	$11^{m}.2$	$9^{m}.9$	$7^{m}.6$
S2	$12^{m}.0$	$10^{m}.8$	$10^{m}.3$
S3	$12^{m}.1$	$12^{m}.0$	$\geq 10^{m}.5$
S4	$\geq 15^{m}.0$	$11^{m}.7$	$6^{m}.4$
π1824	$8^{m}.7$	$8^{m}.7$	$8^{m}.8$

Table 5.5: Photometry of sources in OMC-1S field

H, obvious at K, and the brightest source of all at nbL.

Aperture photometry was performed for each of the four infrared sources, and one of the optical stars, π1824. Although the quality of the data in the H is somewhat poor, with the nbL image even worse, we feel it is worth presenting approximate magnitudes for those sources visible at those wavelengths. The results are given in table 5.5. In all cases the errors are determined by the inaccuracy in subtracting a suitable sky value. At H and K, the errors are roughly $\pm 0^{m}.25$, whilst at nbL, they are as poor as $\pm 0^{m}.5$.

We can see immediately that our photometric procedures are roughly correct, as π1824 has H – K and K – nbL colours of close to zero, consistent with it being an unreddened optically visible star. Of the sources in the cluster of three, S1 and S2 have roughly similar H – K colours of $1^{m}.3$ and $1^{m}.2$, whilst S3 is clearly bluer, with an H – K close to zero. This indicates that S3 may be a faint foreground star in the line of sight towards OMC-1S, or alternatively that it is a reflection nebulosity, associated with S1. In contrast, the K – nbL colours of S1 and S2 are considerably different, at $2^{m}.3$ and $0^{m}.5$ respectively. Although our K – nbL colours are extremely uncertain, due to the errors in the nbL magnitudes, we can still safely assume that S1 is the redder of the two sources in this colour.

The single red source S4 is indeed confirmed to be very red, with an H – K in excess of $3^{m}.3$, and a very red K – nbL of $5^{m}.3$. This is almost $1^{m}.5$ redder than the K – L colour of BN as measured by LBLS.

These 4 infrared sources are therefore candidates for a luminous, deeply embedded source required to supply the observed far infrared luminosity of OMC-1S as discussed

above.

5.4.9.2 Comparison between near infrared and long wavelength positions

One problem that arises when trying decide which of these infrared sources might be responsible for the luminosity of OMC-1S is that the position of the southern molecular cloud core is not well known, due to the limited spatial resolution of the far infrared, sub millimetre, and millimetre maps in which it is seen directly. In order to help resolve this ambiguity, we have collected together all positional measurements of OMC-1S given in the literature. Although the resolution of each individual map is generally about 30″, by plotting all these maps together, we can hope to derive a common centre, which would best locate the core of the molecular cloud. Of course, the obvious assumption that we make in doing so is that all such the relevant emission lines and continuum fluxes peak at a common position. Whilst this may not be exactly the case, we have been careful to choose the lines that are most likely to indicate the position of a dense, cool molecular cloud core, as discussed in the review earlier.

We have found nine far infrared, sub millimetre, and millimetre measurements of OMC-1S in the literature, as reviewed at the beginning of this section. We have obtained positions from each. In order to compensate for possible errors in the absolute astrometric frame of reference used in each case, we have also noted the position given for another source in the map, usually IRc2 or BN, and in one case, θ^1 Ori C. We then calculated the position of OMC-1S in terms of an offset from this other source. As the positions of IRc2, BN, and θ^1 Ori C are accurately determined in the IRCAM K mosaic, we could then calculate the position of OMC-1S relative to these sources in the K mosaic. This procedure rarely changed the original quoted astrometric position of OMC-1S by more than a few arcseconds, but was worth doing for the sake of consistency. The resulting normalised positions for OMC-1S are given in table 5.6, along with the line or wavelength the measurement was made at, and the effective beam size quoted for each map. Note that for the recent measurement of an SiO outflow in the region, we have used the position quoted for an observed symmetric profile, rather than position at which the peak intensity was seen, it having a red shifted wing.

Over the relevant subsection of the IRCAM K mosaic, we have plotted circles with

Line/λ	OMC-1S position (1950.0)		Beam size	Ref.
	RA	Dec	($\hat{''}$)	
400 μm	05 32 46.19	−05 25 51.5	35	[75]
^{13}CO	05 32 45.34	−05 25 57.1	60	[125]
Far-IR	05 32 46.62	−05 25 58.0	80	[36]
H_2CO	05 32 45.24	−05 25 48.0	30	[72]
NH_3	05 32 46.02	−05 26 03.0	40	[16]
NH_3	05 32 45.62	−05 26 06.0	38	[93]
3 mm	05 32 46.15	−05 26 01.3	7.5	[114]
SiO	05 32 46.82	−05 25 59.0	30	[169]
CS J=2–1	05 32 46.12	−05 25 57.0	20	[61]

Table 5.6: Molecular line and continuum positions for OMC-1S

the same diameter as the beam size quoted for each measurement, at the normalised position as given in table 5.6. The resulting image is shown in figure 5.23. For each measurement, the relevant emission line or wavelength is given at about the five o'clock position, just outside the circle, apart from the measurements made at 3 mm and in the J=2–1 line of CS, which have been labelled at ten and three o'clock respectively.

It is immediately clear that it is the cluster of three sources, S1–3, that are the closest K point sources to the common centre of all these measurements. In fact, S2 lies within the small region common to all the measurements, excepting the high resolution measurement made in the continuum at 3 mm, which was made with an interferometer, and has an effective beam size of only 7.5 $\hat{''}$. What is also certainly clear is that S4, the heavily reddened source, lies some way from the centroid. Taking a simple unweighted least squares fit to the positions of the nine plotted far infrared, sub millimetre, and millimetre measurements, we find that the common centre lies within 4.5 $\hat{''}$ of S2, with most of the offset in the southerly direction. The error on the least squares fit is $\pm 9\,\hat{''}$ however, so within the errors, sources S1–3 lie along the line of sight towards OMC-1S, whilst S4 is about 25 $\hat{''}$ away from the common centre, almost 3σ.

In order to reinforce this positional coincidence, we also show some data recently obtained on the newly commissioned James Clerk Maxwell telescope, on Mauna Kea.

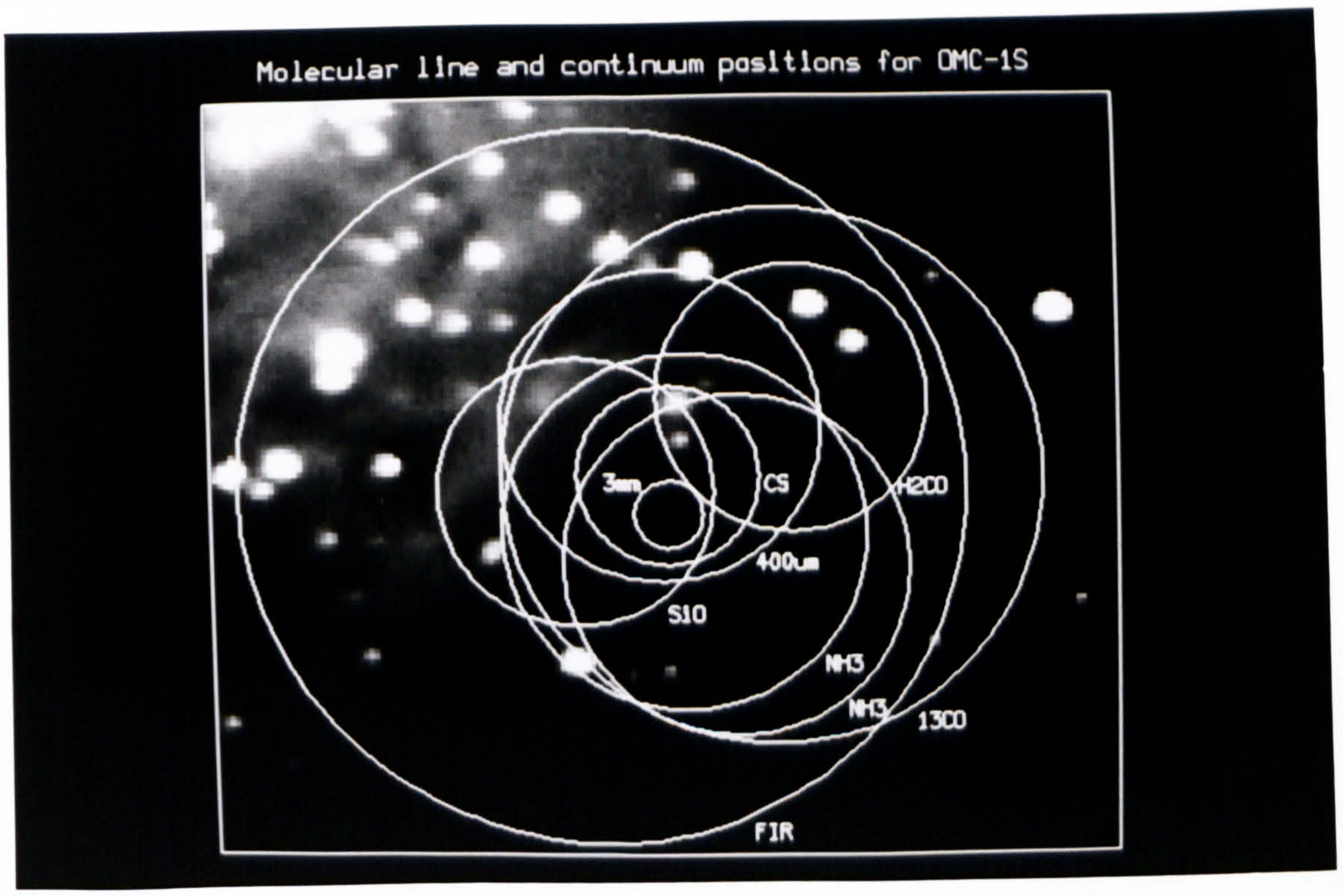


Figure 5.23: Molecular line and continuum positions for OMC-1S

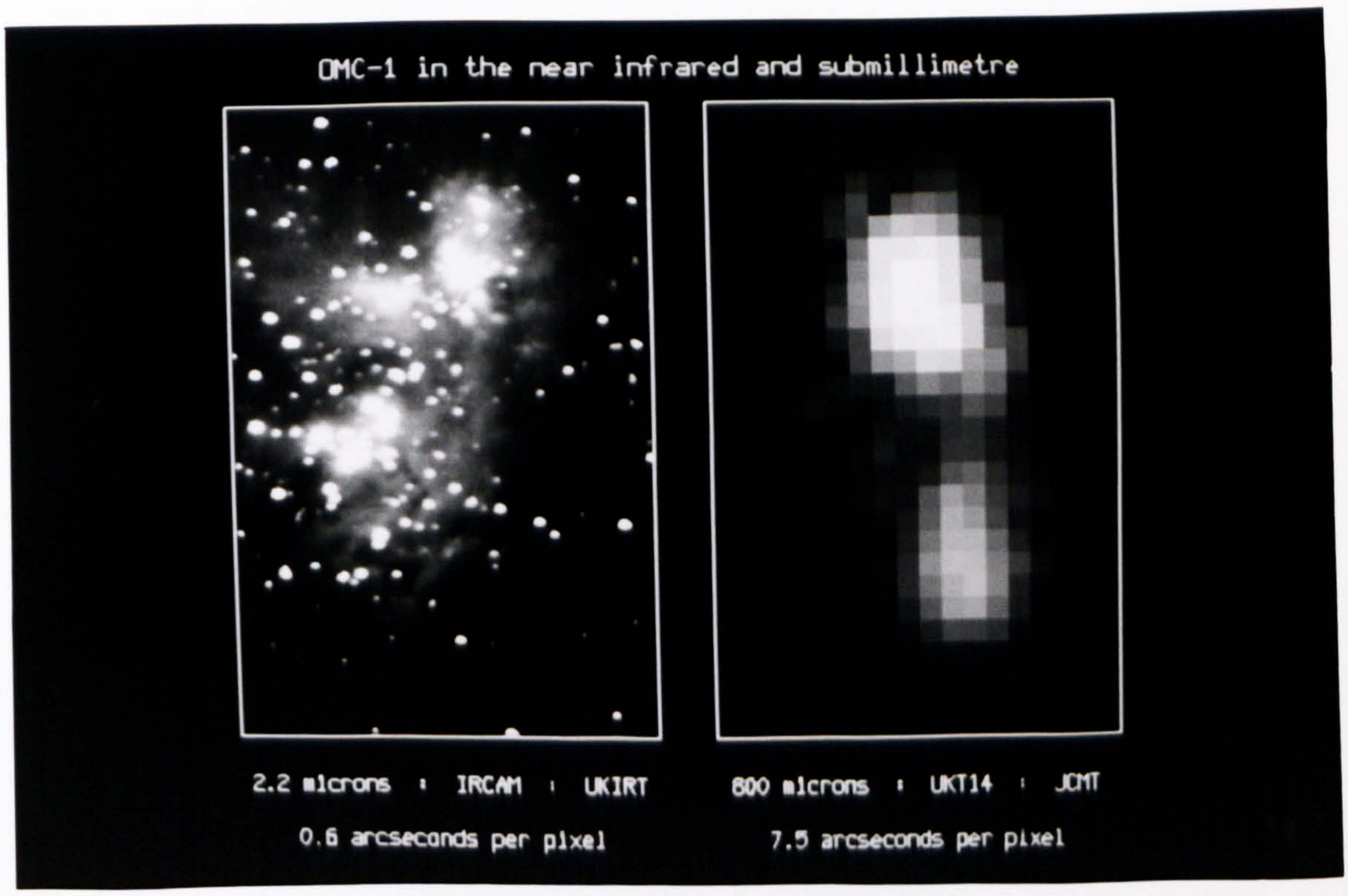


Figure 5.24: 2 μm and 800 μm images of OMC-1

The measurements were made by Göran Sandell and collaborators at a continuum wavelength of 800 μm, using the 15 metre JCMT and the sub millimetre photometer UKT14 to obtain a diffraction limited resolution of 16″, with measurements made every 7.5″. The map covers a 135″ × 210″ region centred between OMC-1N and OMC-1S, and we show it to scale and positionally registered at IRc2 (taken to be at the peak of OMC-1N at 800 μm) alongside a similar sized section of the K mosaic in figure 5.24.

Again, note the coincidence between the location of the secondary peak of the long wavelength emission, and the location of the near infrared point sources S1–3. This image is only shown for reasons of morphological comparison however, as the 800 μm map has yet to be properly corrected for variable atmospheric opacity, and yet to be flux calibrated. An interesting side issue is the 800 μm map pixel size of 7.5″, which corresponds exactly to the beam size used by Becklin *et al.* [18] when they mapped this region at 2 μm with a single channel photometer in the mid-1970s. The comparative resolutions of the 800 μm map and the IRCAM image therefore graphically illustrate the progress made in infrared detector technology over the last decade.

Thus, if our assumption that all the measurements plotted should indicate the position of the cloud core, then we can state that it is the cluster of sources S1–3 that lie along the line of sight towards the core, and not S4. However, purely on the basis of this positional evidence, it would be premature to assume that S1–3 actually lie at the core of OMC-1S, and provide its luminosity.

We do not have sufficient data to unambiguously identify the nature of these sources at this time. By assuming normal type stars, dereddening the data, and deriving absolute magnitudes for the sources, it becomes obvious that a range of possibilities exists which might explain the total observed luminosity of OMC-1S of $\sim 8000\,L_{\odot}$. We may be seeing young pre-main sequence stars which have collapsed rapidly, and have a high luminosity even though relatively cool and red. Alternatively, we may be seeing a main sequence star near B5 type, heavily reddened by extinction in the cloud. The possibility of circumstellar shells and disks surrounding the sources further complicates the picture. Rather than speculate on the basis of very limited data, we can only stress that spectroscopic data are required in order to identify the nature of these sources. Also, we shall be obtaining more detailed imaging observations of this region at J, H, K, and nbL in the near future, and in combination with spectroscopic data, we should

hope to further examine the rôle of these infrared sources with relation to the southern molecular condensation, OMC-1S.

5.4.10 Herbig Haro objects in the Orion Nebula

Although fairly common elsewhere in the sky, few Herbig-Haro (HH) objects have been discovered in HII regions, due to the strong and spatially variable line emission, which makes their discovery difficult via the traditional methods of broad band photography and objective prism surveys. However, as HH objects are thought to be closely related to young stars and the early stages of the stellar evolution process, we would expect to find many in HII regions, which are associated with exactly the same phenomena.

Until recently, only one HH object (M42-HH1) had been identified in the Orion Nebula, using narrow band images of the region in the neutral oxygen line [OI] λ6300Å [112]. In 1980, Canto *et al.* identified another three HH-like objects, using narrow band images in the [NII] λ6584Å and [SII] λ6731Å lines [26]. Two of these sources (M42-HH3 and HH4) were previously known, and are diffuse nebulosities near to θ^2 Ori A. In all four of the HH objects known at this stage, unseen local luminosity sources were thought likely, possibly providing energy via a stellar wind focussed in an elliptical cavity [26].

Subsequent narrow band imaging and velocity resolved spectral imaging in the [OI] λ6300Å line by Axon and Taylor [12], and Taylor *et al.* [138], have revealed a collection of highly blue shifted clumps of material, concentrated in a triangular region to the north-west of the BN-KL region. These observations have led to the suggestion that this collection of HH-like sources is associated with the mass outflow from IRc2. One model uses a precessing jet from IRc2 which illuminates the clumps in turn, whilst a second has the clumps at the interface between the foreground HII region and the background molecular cloud, with photo-ionisation of the clumps by the UV flux from the Trapezium OB cluster and disruption due to shocks from the stellar wind of IRc2 providing the observed line ratios and velocities. In both models, the north-western lobe of the IRc2 bipolar outflow is tilted well into the line of sight, in order to account for the systematically blue shifted line profiles [138].

Thus, recent studies of HH-like objects in Orion have indicated that several different excitation processes may be in operation, from unseen local sources with focussed stellar

winds, to a central illumination centre such as IRc2 providing the energy for a whole group of clumps. The new IRCAM images do not cover the region containing the collection of highly blue shifted clumps, but all four of the previously known HH objects are seen, and are discussed below.

5.4.10.1 M42-HH1

The first discovered HH object in Orion, M42-HH1 is seen to have several high negative velocity components in its emission line spectrum, up to $-288\,\mathrm{kms}^{-1}$ [26]. As it is located about $50''$ to the north-west of BN-KL, Axon and Taylor included it in their collection of sources all illuminated in some way by the stellar wind of IRc2 [12].

However, on discovering it, Münch and Taylor described it as looking almost like a double star, with a separation of about $4''$ in their [OI] images [112]. Although M42-HH1 does look extended on their images in the same line, Axon and Taylor felt there was no real evidence for it being double [12].

In the IRCAM K mosaic, there is a point source near the nominal location of M42-HH1, also visible in the H mosaic, and faintly seen on the AAT I plate. Interestingly, there is also a faint patch of nebulosity just visible in the K image, about $5''$ to the west-north-west of the point source. The point source and nebulosity are referred to as S9 and S10 respectively in the astrometric table in Appendix A. Thus, the point source and patch of nebulosity lie the right distance apart, with the point source lying between BN-KL and the patch of nebulosity, appearing to confirm Münch and Taylor's suggestion of a double source. Although the evidence presented here is very limited indeed, it is possible that we are seeing a continuum point source responsible for the excitation of the more diffuse line emission, perhaps visible here through the K filter in the v=1–0 S(1) line of molecular hydrogen. Obviously, detailed infrared spectroscopy of these sources is required in order to confirm or deny this hypothesis.

5.4.10.2 M42-HH2

In optical narrow band images, M42-HH2 appears as a semi-circular nebulosity at the end of dense neutral filament that runs in a north-westerly direction past the Trapezium

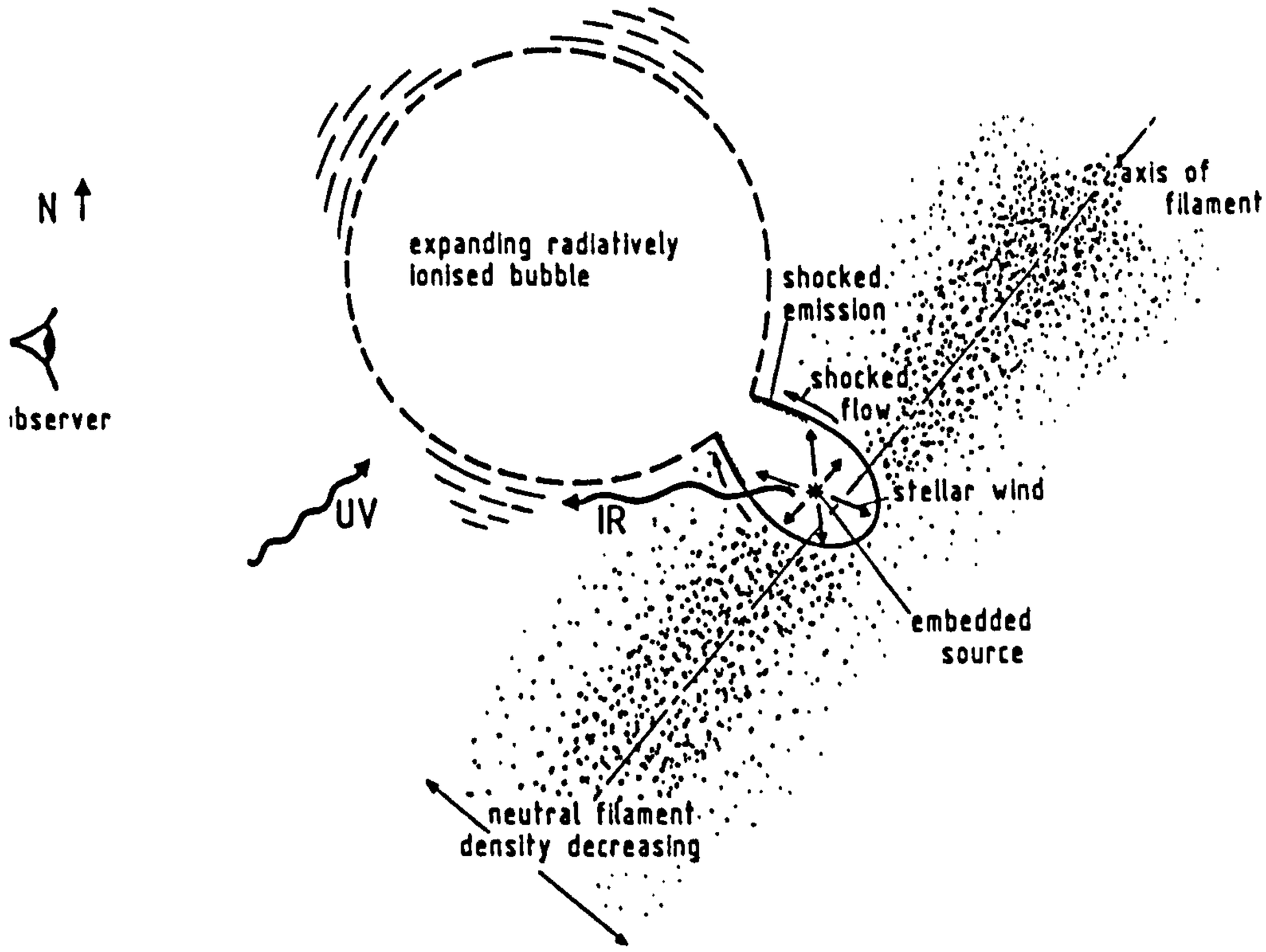


Figure 5.25: Schematic model for M42-HH2

stars [26]. A bright, lowly ionised knot in the semi-circle has been measured by Meaburn as having a velocity of $-54\,\mathrm{kms}^{-1}$ in [SII] and [NII] relative to the general nebular background, and it is surrounded by a more extensive expanding shell with a similar velocity, seen in the [OIII] λ5007Å line [100]. Based on the focussed stellar wind model of Canto *et al.* [26], Meaburn has suggested a model, shown schematically in figure 5.25.

A source embedded in a dense neutral neutral filament, tilted away from the plane perpendicular to the line of sight, creates a cavity with its stellar wind, shocking the edges. This cavity expands preferentially in the direction of decreasing density, and finally creates a hole as it breaks out of the filament. At this point, the streaming stellar wind is focussed at the mouth of the cavity, shocking it, and causing it to be seen in low ionisation lines. Due to extinction between the cavity and the observer, the central source and half of the cavity mouth would be obscured in the optical, unless the neutral filament was lying exactly in the plane perpendicular to the line of sight. On creating a hole in the filament, the stellar wind creates a pressure driven bubble out into the HII region, where it is radiatively ionised by the UV flux of the Trapezium OB stars, giving rise to the more spatially extended higher ionisation emission. The mechanism

by which this bubble is created is acknowledged by Meaburn to be somewhat uncertain.

In the IRCAM images, the semi-circular nebulosity is clearly seen, with a position angle from −30 to 150°. At a p.a. of about 60°, a small and almost pointlike source is seen, whilst at about p.a. 140–150°, a more diffuse knot is seen, which we identify with the optical knot referred to as M42-HH2 by Meaburn. We have numbered these sources as S7 and S8 respectively. However, the dramatic difference between the optical and infrared images is the presence of a fairly bright point source ($K = 9^m.7$) located precisely at the focus of the semi-circular nebulosity. This source is numbered S6. The source is seen at H and K, but not in the optical, and its $H - K$ of $0^m.9$ shows it to be a slightly reddened source.

The immediate suggestion might be that this source S6 is the obscured local illuminating source for the HH-object in Meaburn's model. However, it should be noted that another infrared point source (S9) is seen some $10''$ away from S6, along the line of the optical dense neutral filament towards θ^1 Ori C. It is slightly fainter at $K = 10^m.6$, but with a similar $H - K$ of $0^m.8$, and also unseen in the optical. Despite being further away, this source is more favourably placed relative to the semi-circular diffuse emission in the Meaburn model. This is because the illuminating source should be seen at the focus of the semi-circle only if the axis of the neutral filament were perpendicular to the line of sight. In such a case, we might expect to see a complete circle of emission with the point source at the centre. But as we would be looking straight into the mouth of the cavity, there would be little obscuration towards the point source, and we should see it in the optical. Tilting the axis of the filament away from the plane perpendicular to the line of sight, as shown in the model, would account for the obscuration and reddened $H - K$ colour, and would obscure one half of the cavity mouth, but would also move the source from the focus of the semi-circle, as seen projected on the sky. This fits with S9 being located further south-east along the filament, and with S6 being a chance line of sight alignment.

Clearly however, a more detailed examination of this region is required before any such conclusions can be drawn. Although the brighter of the two new infrared sources, S6, is the more immediately tempting candidate for the rôle as illumination centre of the semi-circular nebulosity containing M42-HH2, the fainter source S9 should also be considered, on geometric grounds.

5.4.10.3 M42-HH3 and HH4 and the nebulosity near θ^2 Ori A

Münch and Wilson first noted the presence of several knots and associated nebulosity close to θ^2 Ori A, just to the south-east of the bright bar [113]. These knots are seen to be bright in the optical emission lines of [OI], [SII], and Hα [54], and radial velocities for three of these features (A, B, C) show high velocity line splitting, indicating motions of up to $\sim 60\,\mathrm{kms}^{-1}$ relative to the underlying nebulosity [139]. These velocities, and their appearance as bow-shocked nebulosities, limb brightened on side facing θ^2 Ori A, led to the suggestion that the knots may be dense cloudlets shocked and accelerated by the stellar wind from θ^2 Ori A [139]. However, it is difficult at accelerate isolated clumps in this way, unless they are small sections of a roughly spherical expanding shell of mainly neutral material, or unless the stellar wind from θ^2 Ori A is anisotropic [26]. Also, the stellar wind from θ^2 Ori A is roughly a factor of ten too small to account for the observed velocities [153]. These problems have led to an alternative suggestion that knots A and B may be Herbig-Haro objects (named M42-HH3 and M42-HH4 [26]), driven by the stellar wind of a newly formed star, embedded in and obscured by the dense neutral material located just to the south-east of the bright bar [26]. A detailed study of the optical [SII] emission line profiles in this region by Walsh [153] results in the same conclusion, with the knots not directly associated with θ^2 Ori A, but more likely driven by a local source. However, Walsh notes that no 2 μm point sources were seen at this location in the low resolution of map Becklin *et al.* [18], and deduces that any such local source must be very heavily embedded in the dense material that lies to the south-east of the bar, just beyond the optical ionisation front [18].

Additional emission line knots have been noted in this region : knot E, approximately $1\hat{}$ to the north-east of θ^2 Ori A [153], and knot F, the third faint component of Walsh's emission line 'triangle', with knots A and B forming the other two vertices. The complex structure linking these three sources is particularly well seen in the light of Hα, [NII], and [OIII] — the famous Lick photograph of Orion in these lines is an excellent example. It is worth noting that Walsh's identification of knot D, just to the north of θ^2 Ori A, is spurious, as this 'knot' is clearly acknowledged (albeit in French) as an optical reflection of θ^2 Ori A in the original 'discovery' paper [85].

All the real knots are visible in the IRCAM K mosaic, and knots A, B, C, and F are indicated in figure 5.26. Knot C is completely dominated by the known associated

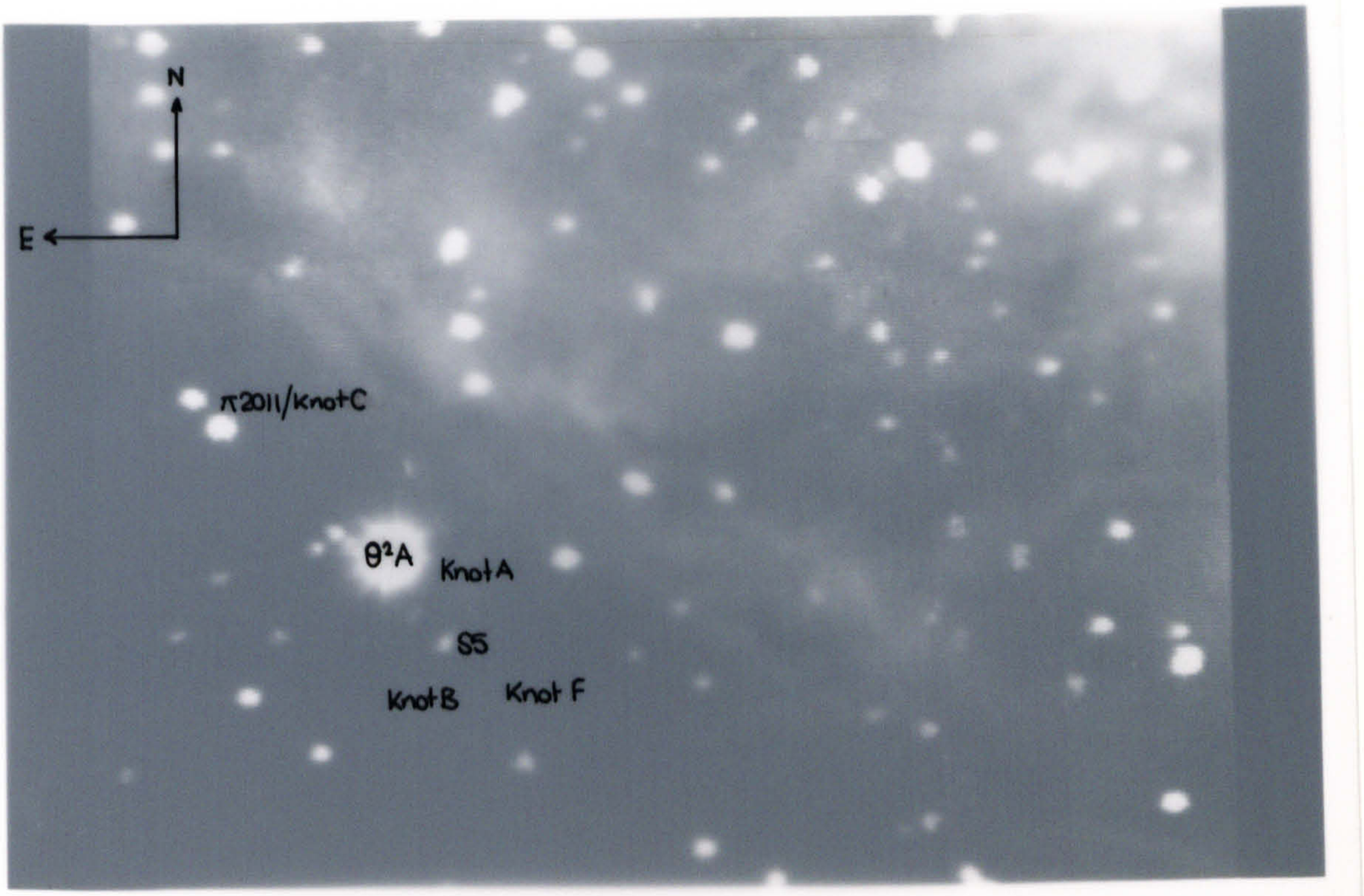
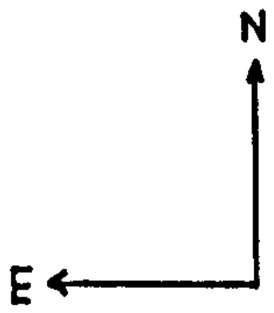


Figure 5.26: θ^2 Ori A and the bright bar

Knot	Peak surface brightness ($\times 10^{-20}$ W/cm^2/$\hat{\pi}^2$)
A	3.2 ± 0.5
B	3.4 ± 0.4
E	3.2 ± 0.4
F	1.0 ± 0.6

Table 5.7: Extended emission nebulosity near θ^2 Ori A

optical star, π2011, which has a V magnitude of $13^{\rm m}.8$ [139], and is seen here at K = $9^{\rm m}.05$. The other knots are nebulous, and their approximate peak surface brightnesses are listed in table 5.7. Unfortunately, the H mosaic does not extend far enough to cover this region. However, the most notable difference between the optical emission line images and the IRCAM K mosaic is the presence in the IRCAM image of a point source (K $\sim 10^{\rm m}.5$) located between knots A and B (M42 HH-3 and HH-4). Due to its proximity to θ^2 Ori A, it was thought that this might be a ghost reflection of the bright star, but its existence is confirmed on two of the separate IRCAM images that make up this part of the mosaic, with θ^2 Ori A in different locations in each image. Also, this point source is clearly seen on the AAT I plate. We have labelled this source as S5.



π2011/knotC

θ²A KnotA

S5

KnotB KnotF

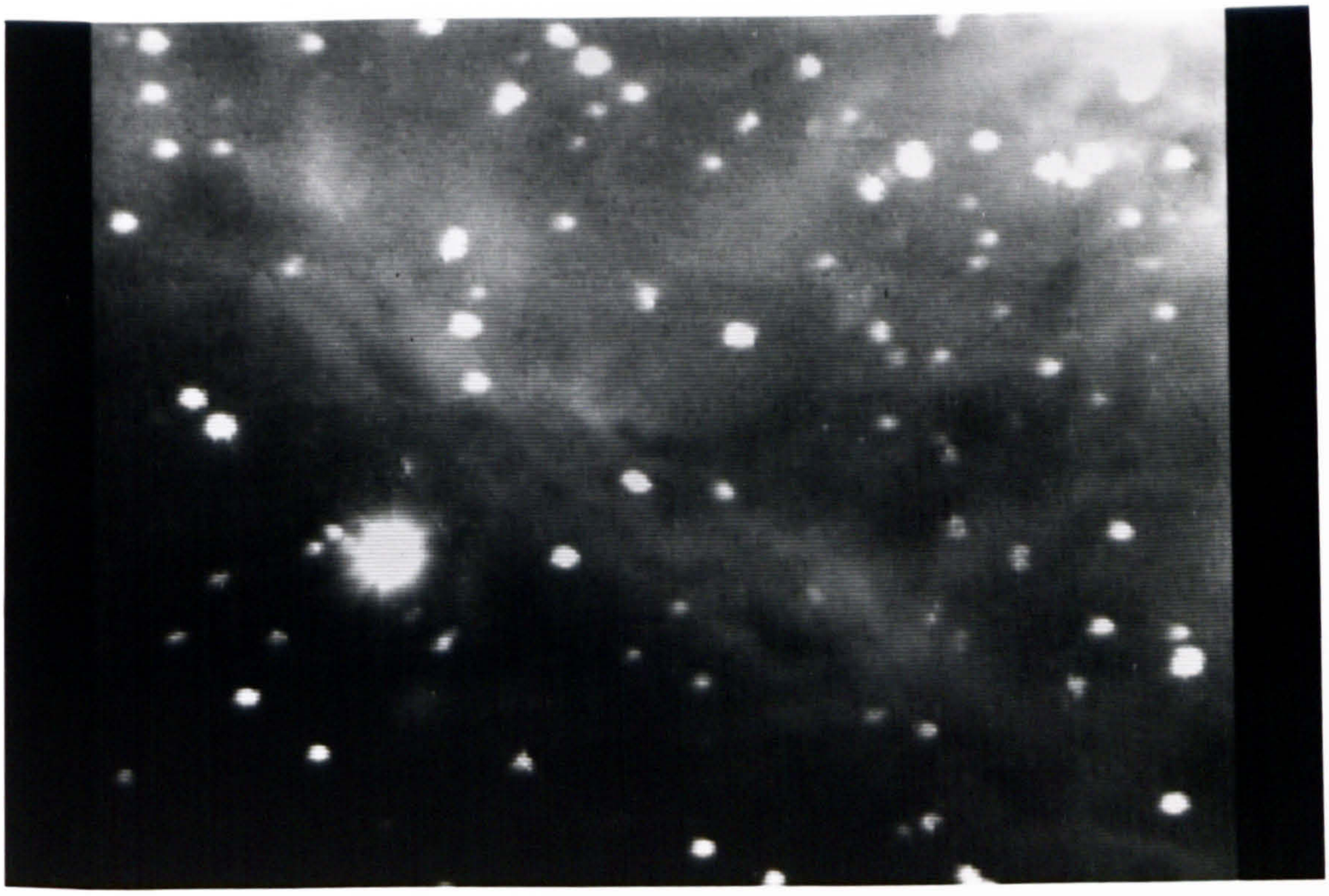


Figure 5.26: θ^2 Ori A and the bright bar

Knot	Peak surface brightness ($\times 10^{-20}$ W/cm^2/$\widehat{\text{n}}^2$)
A	3.2 ± 0.5
B	3.4 ± 0.4
E	3.2 ± 0.4
F	1.0 ± 0.6

Table 5.7: Extended emission nebulosity near θ^2 Ori A

optical star, π2011, which has a V magnitude of $13^{\text{m}}.8$ [139], and is seen here at K = $9^{\text{m}}.05$. The other knots are nebulous, and their approximate peak surface brightnesses are listed in table 5.7. Unfortunately, the H mosaic does not extend far enough to cover this region. However, the most notable difference between the optical emission line images and the IRCAM K mosaic is the presence in the IRCAM image of a point source (K $\sim 10^{\text{m}}.5$) located between knots A and B (M42 HH-3 and HH-4). Due to its proximity to θ^2 Ori A, it was thought that this might be a ghost reflection of the bright star, but its existence is confirmed on two of the separate IRCAM images that make up this part of the mosaic, with θ^2 Ori A in different locations in each image. Also, this point source is clearly seen on the AAT I plate. We have labelled this source as S5.

Thus, as proposed by Canto *et al.* [26] and Walsh [153], S5 is probably an embedded star, and may well be responsible for the excitation of knots A and B. Infrared spectroscopy and imaging polarimetry of these knots and the intervening point source will be necessary in order to determine whether the knots are intrinsic emission line nebulosities, or reflection nebulosities, reflecting the emission of the embedded central source. Knot F is probably associated with knots A and B in some way [153], and although its location perpendicular to the axis joining A, B, and the point source, is difficult to explain in the context of a classical bipolar reflection nebula, it may still be related to the point source in some way.

To summarise : knots A, B, and possibly F, are likely related to the embedded point source; knot C appears to be associated with the star π2011; knot E is a large distance ($\sim 1\,\hat{'}$) to the north-east of θ^2 Ori A; and knot D is non-existent. This virtually eliminates the need to invoke θ^2 Ori A as the major source of excitation for any of the nebulosity around it.

5.4.10.4 Summary of Herbig-Haro object observations

On locating four of the Herbig-Haro objects in the Orion Nebula in our infrared images, we have found probable localised illumination candidates for each. As the collection of HH-like objects to the north-west, unobserved in our images, seems to be best explained in terms of a single illumination centre, IRc2, it is likely that two or more mechanisms for the excitation of HH-type sources are in action in the Orion Nebula. As stated at several points, infrared spectroscopy is urgently required in order to give a more physical basis to the arguments given here, which are largely based on morphological grounds.

5.4.11 The bright bar

Approximately two arcminutes to the south-east of the Trapezium OB stars, the bright bar is an ionisation front running from north-east to south-west, and is clearly visible in the K mosaic as a highly structured ridge of emission. A close up view can be seen in figure 5.26. It is encroaching into a foreground fold of the molecular cloud which lies behind the Trapezium, and is viewed almost edge on [52, p. 95, fig. 2.1.12]. Thermal emission from the dense molecular fold is seen at 10 μm, lying just beyond the optical

ionisation front [18]. At optical wavelengths, the bar is conspicuous in a variety of emission lines, including [OI] (λ6300Å), [OII] (λ3726Å), [SIII] (λ9069Å), and so on [52, chap. 2]. Particularly striking however, is the correspondence between the K image and photographs taken in the light of Hα (λ6563Å) and [NII] (λ6584Å) (for example, see [113,139]). On the Trapezium side of the centre of the bright bar, the 2μm and Hα structures match exactly. On the θ^2 Ori A side of the centre however, nebulosity is still seen at K considerably further south-east than it is seen in the optical. If the Hα cut-off delineates the edge of the ionisation region, then 2μm nebulosity seen to the south-east of that cut-off may be due to molecular hydrogen, shocked at the interface between the neutral HI and molecular regions, beyond the HI–HII interface.

Low spatial resolution 2μm spectroscopy in a strip perpendicular to the bar [62] shows that the peak H_2 v=1–0 S(1) flux is seen some 30$\widehat{''}$ to the south-east of the peak Brγ position, consistent with the picture drawn above. However, at no point in the strip does the molecular hydrogen flux exceed that from the ionised hydrogen, and Brγ flux is still seen well to the south-east of the Hα cut-off, making this simple picture somewhat less clear. Additionally, the ratio of the v=1–0 S(1) and v=2–1 S(1) line fluxes decreases to the south-east of the bar, suggesting an increase in radiatively pumped H_2 emission, as opposed to shock excited emission [62]. The pumping source of this fluorescent H_2 emission may be θ^2 Ori A, and this will further complicate an simple picture of the ionisation ridge profile.

The peak flux observed in the bright bar in the IRCAM K mosaic is $\sim$ 15 $\times$ 10^{-20} W/cm^2/$\widehat{''}^2$, with a mean value 2–3 times smaller. This mean value is 10–20 times greater than the flux observed per square arcsecond in the Brγ line, when averaged across a 19.6$\widehat{''}$ diameter beam [62]. Inspection of a low resolution CVF spectrum of the whole K window shows that the integrated continuum + Brγ + H_2 flux from the bar is about 10 times higher than the Brγ flux alone.

Thus, the flux observed in the bright bar in the K image is consistent with that seen in large beam spectroscopy integrated across the whole K window. The majority of the emission nebulosity seen in the bright bar in the K image is due to continuum, rather than ionised or shocked hydrogen emission lines. Nevertheless, the strong correspondence between the Hα and K images in many bright knots and dark regions of the bar suggests that a Brγ : continuum ratio much larger than 1 : 10 might exist on scales

of a few arcseconds, too small to be probed by the previous large beam spectroscopy. Alternatively, the dark regions may simply be foreground extinction towards the bar, obscuring both emission line nebulosity, and continuum emission or reflected flux in similar ways.

5.4.12 Compact radio sources

In addition to the extended radio emission associated with the ionised gas of the M42 HII region, recent observations at high spatial resolution with the VLA have revealed a very high density of very compact ($0.1^{\prime\prime}$–$0.2^{\prime\prime}$) radio sources within the central $3^{\prime} \times 3^{\prime}$ core. These studies have been carried out by Garay *et al.* [45], Churchwell *et al.* [29], and Garay [44], who summarises the classification of the sources as follows :

- Twenty five sources are clustered near the Trapezium stars, and are thought to be dense neutral condensations surrounded by ionised envelopes, excited by the brightest Trapezium star, θ^1 Ori C — these are referred to as PIGS, for partially ionised globules. An alternative explanation for these sources is that they may be evaporating protostellar disks of $\sim$ solar mass [29]. Many have visible counterparts.

- Four sources are located near the BN-KL cluster and none have optically visible counterparts. Two are coincident with the infrared sources IRc2 and BN, and hence they are thought to be recently formed luminous stars in the molecular cloud — they are known as DEERS, for deeply embedded energetic radio sources.

- The last four sources are in dark bay, to the east of BN-KL — these are known to have optical counterparts, fluctuate, and emit X-rays, hence the acronym FOXES, for fluctuating optical and X-ray emitting sources. They may well be T Tauri stars.

As Garay notes however, these conclusions as to the nature of the compact radio sources are tentative, and further observations are required, including optical and infrared spectroscopy of the PIGS and FOXES, and infrared photometry and high spatial resolution sub-millimetre observations of the embedded DEERS. The first question however is whether the compact radio sources are associated with stars, visible or obscured, or nebulous condensations. Seven of the PIGS are associated with the nebulous optical

emission line condensations noted by Laques and Vidal [85], and eighteen of the remaining sources coincide with optical stars seen on far red photographs [45]. As we have seen, infrared observations allow greater penetration of the dust and nebulosity seen in the Orion Nebula. To this end, Garay *et al.* [45] and Churchwell *et al.* [29] have used the 2 μm map of LBLS to check for infrared identifications, whilst Garay has used the multicolour images of ABH. Twenty seven of the compact radio sources are identified in the 2$^{\prime\prime}$ resolution image of Allen *et al.* , but several others lie outside the boundaries of their image. All the sources listed by Garay lie within the boundaries of the IRCAM K mosaic, and in addition, the higher spatial resolution of our K image should allow us to better determine the spatial extent of any sources detected.

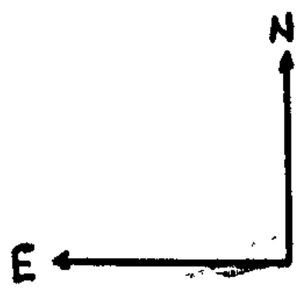
All but three of the thirty three compact radio source listed by Garay are visible in the IRCAM K mosaic. Of the remaining three, two are the DEERS D and H, which lie in the direction of the strip of extinction that runs from south-west to north-east across IRc2 — these sources may be heavily obscured high luminosity sources, or low luminosity radio emitting young objects [44], but this cannot be decided in the absence of any information other than their radio emission properties. If they are luminous embedded sources, they may be visible at L′ and M. The other source lies in the wing of θ^1 Ori B, and is impossible to detect due to the brightness of this star. A summary of the thirty sources identified in the IRCAM image are given in table 5.8, along with their VLA and IRCAM positions. We have indicated the positions of some of the PIGS in a subsection of the large K mosaic, centred on the Trapezium region — see figure 5.27. As the positional accuracy for the compact sources measured by the VLA is $\sim 0.1^{\prime\prime}$ [45], we can compare the VLA positions with the IRCAM positions as a secondary check of the astrometric accuracy of the IRCAM mosaic, as discussed in section 5.4.3. The RMS deviation between the VLA and IRCAM positions for the thirty common sources is calculated as $0.6^{\prime\prime}$, which is consistent with the accuracy derived for the mosaic from the optical astrometry.

5.4.13 Conclusions

We have examined our new IRCAM mosaic of the Orion Nebula at K in detail. The data compare very favourably to that published prior to the availability of infrared arrays.

No.	Positions (1950.0)				Offsets		Other names	Note
	VLA		IRCAM		RA	Dec		
PIGS								
1	5 32 50.21	−5 25 34.1	5 32 50.22	−5 25 34.8	0.12	0.71		Point
2	5 32 50.10	−5 25 18.2	5 32 50.06	−5 25 18.6	0.57	0.40		Extended
3	5 32 49.61	−5 25 27.3	5 32 49.64	−5 25 27.9	0.41	0.62	π1893	Point
4	5 32 49.52	−5 25 30.3	5 32 49.51	−5 25 31.0	0.17	0.72	π1984	Point
5	5 32 49.39	−5 25 19.6	—	—	—	—	LV1	Wing θ^1C
6	5 32 49.29	−5 25 09.8	5 32 49.32	−5 25 09.8	0.39	0.03	π1890,θ^1G,LV2	Point
7	5 32 48.82	−5 25 10.0	5 32 48.81	−5 25 10.4	0.10	0.41	LV3	Point
8	5 32 48.60	−5 25 17.7	5 32 48.60	−5 25 17.9	0.03	0.17	LV4	Wing θ^1C
9	5 32 48.49	−5 25 43.1	5 32 48.54	−5 25 43.4	0.70	0.31	π1869	Point
10	5 32 48.39	−5 25 18.9	—	—	—	—	LV6	Not seen
11	5 32 48.37	−5 25 15.9	5 32 48.39	−5 25 16.0	0.30	0.09	π1867,θ^1H,LV5	Binary
12	5 32 48.35	−5 25 07.5	5 32 48.40	−5 25 07.9	0.69	0.39	π1865,θ^1A	Point
13	5 32 48.33	−5 25 20.0	5 32 48.30	−5 25 20.3	0.41	0.34	LV6	Extended
14	5 32 48.06	−5 25 30.9	5 32 48.04	−5 25 31.5	0.26	0.64		Point
15	5 32 48.60	−5 25 00.5	—	—	—	—		Wing θ^1B
16	5 32 48.87	−5 25 16.0	—	—	—	—		Wing θ^1C
17	5 32 49.31	−5 25 21.3	5 32 49.26	−5 25 21.7	0.68	0.36		Point
19	5 32 50.58	−5 25 24.0	5 32 50.56	−5 25 24.9	0.30	0.87		Faint
20	5 32 49.38	−5 25 19.3	—	—	—	—		Wing θ^1C
21	5 32 49.15	−5 25 09.4	5 32 49.15	−5 25 09.8	0.03	0.42		Point
22	5 32 48.61	−5 25 21.2	5 32 48.60	−5 25 21.6	0.22	0.41		Faint
23	5 32 48.54	−5 25 46.4	5 32 48.58	−5 25 46.5	0.54	0.13	π1870	Point
24	5 32 48.44	−5 25 31.4	5 32 48.42	−5 25 31.6	0.30	0.16		Faint
25	5 32 48.30	−5 25 03.3	5 32 48.32	−5 25 03.5	0.24	0.23	π1864,θ^1E	Point
26	5 32 48.27	−5 25 15.9	5 32 48.26	−5 25 16.6	0.09	0.70	π1866,θ^1H	Binary
DEERS								
B	5 32 46.64	−5 24 16.5	5 32 46.63	−5 24 16.7	0.08	0.18	BN	Point
H	5 32 47.02	−5 24 32.1	—	—	—	—		Not seen
I	5 32 47.02	−5 24 24.0	5 32 47.01	−5 24 23.6	0.21	0.44	IRc2	Faint
D	5 32 47.41	−5 24 19.0	—	—	—	—		Not seen
FOXES								
E	5 32 49.50	−5 24 41.9	5 32 49.50	−5 24 42.4	0.06	0.54		Faint
L	5 32 49.87	−5 24 29.3	5 32 49.85	−5 24 30.0	0.33	0.70	π1909	Point
G	5 32 50.47	−5 24 38.7	5 32 50.47	−5 24 39.4	0.01	0.69	π1910	Binary
F	5 32 50.89	−5 24 30.6	5 32 50.89	−5 24 30.7	0.06	0.10	π1925	Point

Table 5.8: Identifications of compact radio sources



25

12

6 21 7

8 11 26

2

22 13

17

3

4

24 14

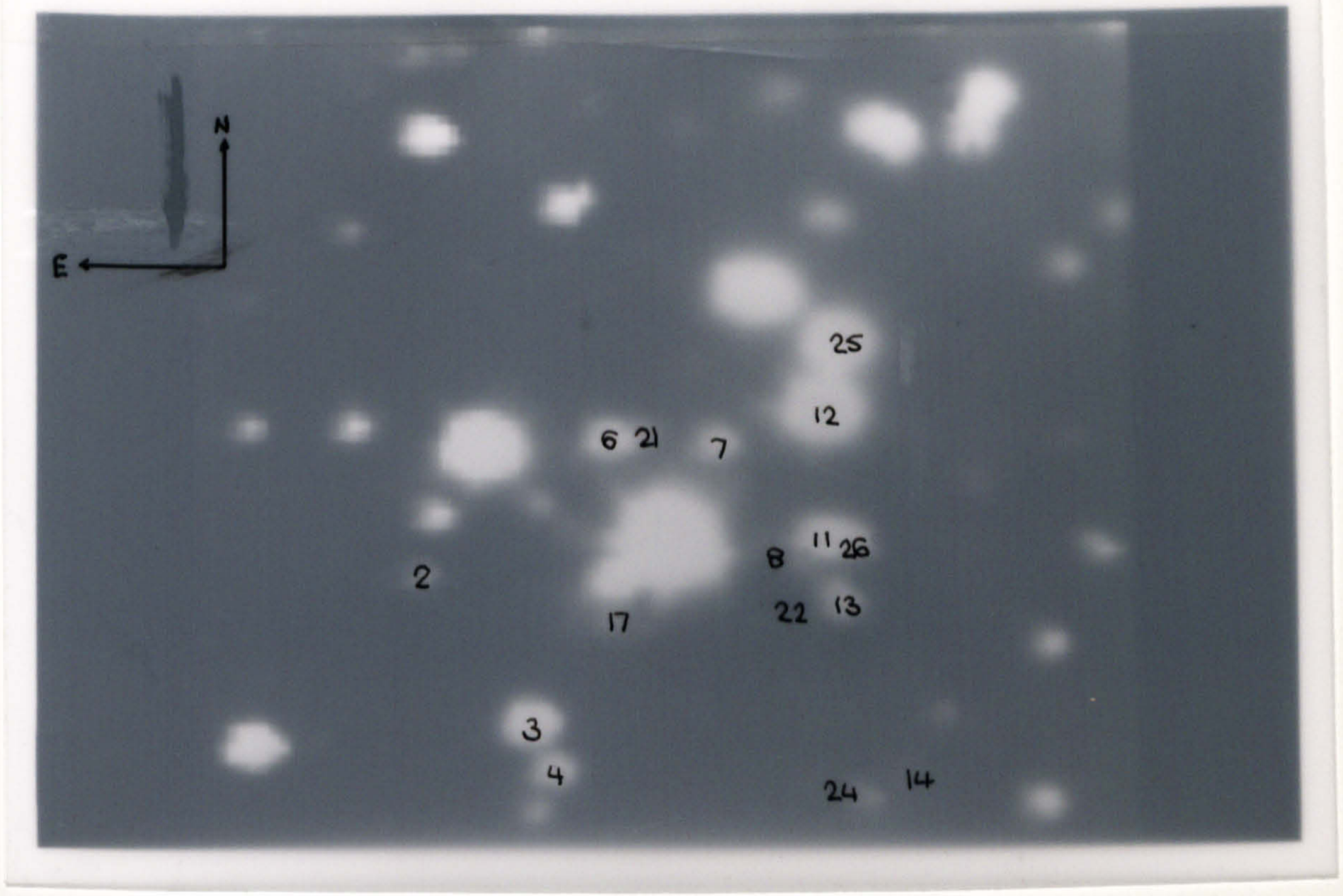


Figure 5.27: The Trapezium and surrounding compact radio sources

We have been largely restricted to morphological comments, due to the lack of complementary wavelength data. Nevertheless, we have seen that considerable information is available even at the morphological level, and we have serendipitously identified several new sources in this work, including embedded stars possibly responsible for the excitation of four known Herbig-Haro-like sources, and four infrared point sources towards the southern molecular condensation, OMC-1S. Limited colour information for the latter sources indicates that two of a cluster of three sources are fairly red, and that an additional source some 15″ to the north is extremely red indeed. Based on comparisons with far infrared, submillimetre, and millimetre data, we have been able to identify which sources lie closest to the line of sight towards the peak emission from OMC-1S.

There is obviously a need for much further work based on the morphological results of this work. Spectroscopy of the newly found sources is needed to identify their nature, whilst large scale multicolour imaging photometry is required to more fully understand the distribution of the large number of faint stars seen throughout the image. This latter work is planned, and should provide results with substantially enhanced sensitivity, when compared to the data presented here, which were taken with an engineering grade

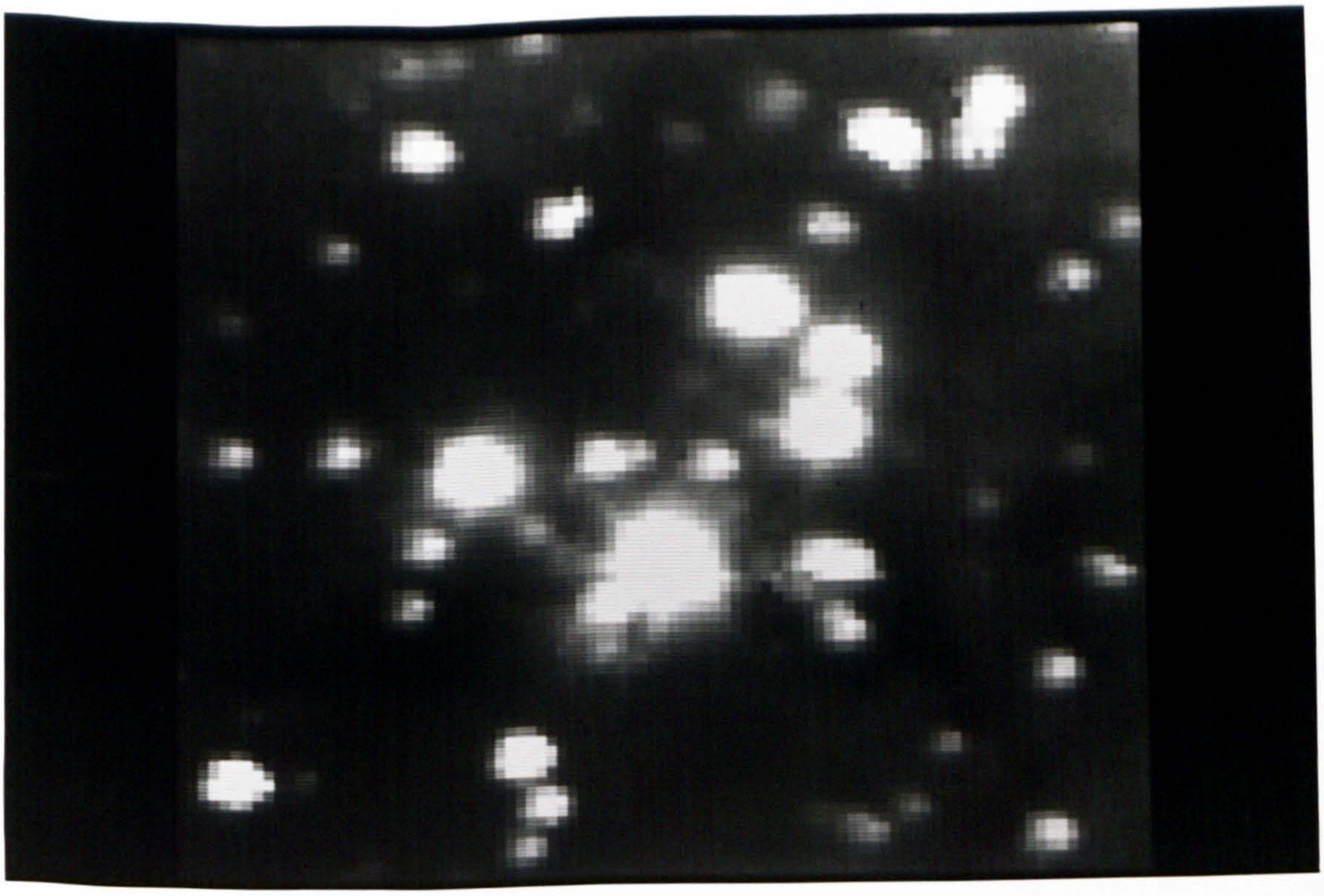


Figure 5.27: The Trapezium and surrounding compact radio sources

We have been largely restricted to morphological comments, due to the lack of complementary wavelength data. Nevertheless, we have seen that considerable information is available even at the morphological level, and we have serendipitously identified several new sources in this work, including embedded stars possibly responsible for the excitation of four known Herbig-Haro-like sources, and four infrared point sources towards the southern molecular condensation, OMC-1S. Limited colour information for the latter sources indicates that two of a cluster of three sources are fairly red, and that an additional source some 15$\widehat{''}$ to the north is extremely red indeed. Based on comparisons with far infrared, submillimetre, and millimetre data, we have been able to identify which sources lie closest to the line of sight towards the peak emission from OMC-1S.

There is obviously a need for much further work based on the morphological results of this work. Spectroscopy of the newly found sources is needed to identify their nature, whilst large scale multicolour imaging photometry is required to more fully understand the distribution of the large number of faint stars seen throughout the image. This latter work is planned, and should provide results with substantially enhanced sensitivity, when compared to the data presented here, which were taken with an engineering grade

array under extremely high dark current and sky background conditions.

However, we have clearly demonstrated the power of infrared arrays to obtain large scale seeing limited images in relatively small amounts of telescope time. As the techniques applied to observing with infrared cameras develop, we are certain to see great improvements in the quality of data available. Also, considerable effort is likely to be invested in optimising data reduction and analysis procedures. Already however, we can see that detector arrays will truly revolutionise infrared astronomy.

Bibliography

[1] *CFHT Observers Manual.* CFH Telescope Corporation, 1983.

[2] Infrared focal plane arrays. Opt. Eng. vol. 26(3), March 1987.

[3] *Infrared Handbook.* Optical Coating Laboratory, Inc., Santa Rosa, CA, 1970.

[4] *KPNO Facilities Book.* Kitt Peak National Observatory, 1987.

[5] *Optics and filters.* Oriel Corporation, Stratford, CT, 1984. Volume III.

[6] D. J. Adams, A. J. Adamson, and A. B. Giles. A 2.2 μm map of NGC5128. *Mon. Not. R. astr. Soc.*, 202:241–248, 1983.

[7] D. A. Allen. An infrared image of the Hourglass region of M8. *Mon. Not. R. astr. Soc.*, 319:35P–38P, 1986.

[8] D. A. Allen, J. A. Bailey, and A. R. Hyland. Infrared images of the Orion Nebula. *Sky and Telescope*, 67:221–224, 1984. (ABH).

[9] D. A. Allen and J. R. Barton. A study of sky noise, 1.5 μm–5.5 μm. *Pub. A. S. P.*, 93:381–384, 1981.

[10] C. Aspin, I. S. McLean, and M. J. McCaughrean. CCD observations of bipolar nebulæ : II LkHα233. *Astron. Astrophys.*, 144:220–226, 1985.

[11] C. Aspin, I. S. McLean, and J. T. Rayner. IR imaging of bipolar nebulæ/outflows. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 281–285, Institute for Astronomy, University of Hawaii, 1987.

[12] D. J. Axon and K. Taylor. Discovery of a family of Herbig-Haro objects in M42 : implications for the geometry of the high velocity molecular flow ? *Mon. Not. R. astr. Soc.*, 207:241–261, 1984.

[13] G. C. Bailey. FET switch multiplexers for large area infrared focal planes. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 52–59, Institute for Astronomy, University of Hawaii, 1987.

[14] J. R. Barton and D. A. Allen. An integrating pre-amplifier for indium antimonide detectors. *Pub. A. S. P.*, 92:368–374, 1980.

[15] P. Bastien, W. Batrla, C. Henkel, T. Pauls, C. M. Walmsley, and T. L. Wilson. Formaldehyde observations of OMC-1. *Astron. Astrophys.*, 146:86–94, 1985.

[16] W. Batrla, T. L. Wilson, P. Bastien, and K. Ruf. Clumping in molecular clouds. The region between OMC1 and 2. *Astron. Astrophys.*, 128:279–290, 1983.

[17] E. Battaner, J. E. Beckman, E. Mediavilla, M. Prieto, C. Sanchez-Magro, C. Mũnoz-Tũnon, and M. L. Sanchez Saavedra. Near-infrared mapping of spiral galaxies II. J, H, K profiles of M31. *Astron. Astrophys.*, 161:70–74, 1984.

[18] E. E. Becklin, S. Beckwith, I. Gatley, K. Matthews, G. Neugebauer, C. Sarazin, and M. W. Werner. Infrared studies of an ionization front region in the Orion Nebula. *Ap. J.*, 207:770–779, 1976.

[19] E. E. Becklin and G. Neugebauer. Observations of an infrared star in the Orion Nebula. *Ap. J.*, 147:799–802, 1967.

[20] S. Beckwith. The implications of molecular hydrogen emission. In C. G. Wynn-Williams and D. P. Cruikshank, editors, *Infrared Astronomy*, pages 167–178, IAU Symposium vol. 90, D. Reidel, 1981.

[21] S. Beckwith, S. E. Persson, G. Neugebauer, and E. E. Becklin. Observations of molecular hydrogen emission from the Orion Nebula. *Ap. J.*, 223:464–470, 1978.

[22] P. Y. Bely. Weather and seeing on Mauna Kea. *Pub. A. S. P.*, 99:560–570, 1987.

[23] M. C. Bird. *Integrating detectors and applications to infrared spectroscopy.* PhD thesis, University of Edinburgh, 1988. (in preparation).

[24] D. J. Bradley and P. N. J. Dennis. Sampling effects in CdHgTe focal plane arrays. In L. R. Baker and A. Masson, editors, *Infrared technology and applications*, pages 53–60, Proc. SPIE vol. 590, 1986.

[25] D. S. Brown. The imaging potential of large groundbased infrared telescopes. In A. Hewitt, editor, *Optical Telescopes for the 1990's*, pages 999–1005, KPNO Tucson, 1980.

[26] J. Cantó, C. Goudis, P. G. Johnson, and J. Meaburn. Herbig-Haro objects in the Orion Nebula. *Astron. Astrophys.*, 85:128–134, 1980.

[27] B. R. Capone, R. W. Taylor, and W. F. Kosonocky. Design and characterization of a Schottky infrared charge coupled device (IRCCD) focal plane array. *Opt. Eng.*, 21:945–950, 1982.

[28] J. W. Chamberlain. *Physics of the Aurora and Airglow.* Volume 2 of *International Geophysics Series*, Academic Press, 1961.

[29] E. Churchwell, M. Felli, D. O. S. Wood, and M. Massi. Solar-system sized condensations in the Orion Nebula. *Ap. J.*, 321:516–529, 1987.

[30] C. E. Coulman. Fundamental and applied aspects of astronomical "seeing". *Ann. Revs. Astron. Astrophys.*, 23:19–57, 1985.

[31] R. J. Dann, S. R. Carpenter, C. Seamer, P. N. J. Dennis, and D. J. Bradley. Sampling effects in CdHgTe focal plane arrays — practical results. In I. J. Spiro and R. A. Mollicone, editors, *Infrared technology XII*, pages 123–128, Proc. SPIE vol. 685, 1986.

[32] K. A. Dick. On the rotational temperature of airglow hydroxyl emission. *Planet. Space Sci.*, 25:595–596, 1977.

[33] F. Diego. Stellar image profiles from linear detectors and the throughput of astronomical instruments. *Pub. A. S. P.*, 97:1209–1214, 1985.

[34] R. D. Dietz, J. Smith, J. A. Hackwell, R. D. Gehrz, and G. L. Grasdalen. Morphology of the nuclear region of M82 at 2.2μ. *Astron. J.*, 91:758–760, 1986.

[35] D. Downes, R. Genzel, Å. Hjalmarson, L. Å. Nyman, and B. Rönnäng. Thermal SiO as a probe of high velocities in regions of star formation. *Ap. J.*, 252:L29–L33, 1982.

[36] S. Drapatz, L. Haser, R. Hofmann, N. Oda, and K. V. K. Iyengar. Far-infared spectrophotometry of the Orion Molecular Cloud 1 ridge. *Astron. Astrophys.*, 128:207–211, 1983.

[37] W. J. Forrest. Ground based imaging with a 32 × 32 InSb array. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 171–179, Institute for Astronomy, University of Hawaii, 1987.

[38] W. J. Forrest, A. Moneti, C. E. Woodward, J. L. Pipher, and A. W. Hoffman. The new near infrared array camera at the University of Rochester. *Pub. A. S. P.*, 97:183–198, 1985.

[39] W. J. Forrest, M. A. Shure, J. L. Pipher, and C. E. Woodward. Brackett alpha images of the Galactic Center. In D. C. Backer, editor, *The Galactic Center — proceedings of the symposium honoring C. H. Townes*, pages 153–156, American Institute of Physics vol. 155, 1987.

[40] A. M. Fowler and J. P. Britt. InSb charge-injection device array performance. In D. L. Crawford, editor, *Instrumentation in Astronomy IV*, pages 18–25, Proc. SPIE vol. 331, 1982.

[41] A. M. Fowler, F. C. Gillett, B. Gregory, R. R. Joyce, R. G. Probst, and R. Smith. The NOAO infrared imagers : description and performance. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 197–203, Institute for Astronomy, University of Hawaii, 1987.

[42] A. M. Fowler, R. G. Probst, J. P. Britt, R. R. Joyce, and F. C. Gillett. Evaluation of an indium antimonide hybrid focal plane array for ground-based astronomy. *Opt. Eng.*, 26:232–240, 1987.

[43] M. Gadsden and P. C. Wraight. Atmospheric transmission of the 1.27 μm band of oxygen. *Jnl. Atmos. Terr. Phys.*, 37:287–296, 1975.

[44] G. Garay. *The Orion radio zoo : PIGS, FOXES and DEERS.* European Southern Observatory, April 1987. preprint no. 497.

[45] G. Garay, J. M. Moran, and M. J. Reid. Compact continuum radio sources in the Orion Nebula. *Ap. J.*, 314:535–550, 1987.

[46] R. Garden, T. R. Geballe, I. Gatley, and D. N. Nadeau. An extremely luminous bipolar H_2 flow in the DR21 star forming region. *Mon. Not. R. astr. Soc.*, 220:203–221, 1986.

[47] R. D. Gehrz. Matching infrared array instruments to future large telescopes. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 499–507, Institute for Astronomy, University of Hawaii, 1987.

[48] R. D. Gehrz, J. A. Hackwell, and J. R. Smith. 8–13 micron maps of the Trapezium region of the Orion Nebula. *Ap. J.*, 202:L33–L36, 1975.

[49] M. D. Gibbons and S. C. Wang. Status of CID InSb detector technology. In W. L. Wolfe, editor, *Infrared Detectors*, pages 151–166, Proc. SPIE vol. 443, 1983.

[50] F. C. Gillett. Infrared arrays for ground-based astronomy. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 3–12, Institute for Astronomy, University of Hawaii, 1987.

[51] F. C. Gillett, E. L. Dereniak, and R. R. Joyce. Detectors for infrared astronomy. *Opt. Eng.*, 16:544–550, 1977.

[52] C. Goudis. *The Orion Complex : A case study of interstellar matter.* Volume 90 of *Astrophysics and Space Science Library*, D. Reidel, 1982.

[53] G. L. Grasdalen, J. A. Hackwell, and R. D. Gehrz. Imaging with a single detector : the Wyoming approach. *Pub. A. S. P.*, 96:1017–1023, 1984.

[54] T. R. Gull. The Orion Nebula : a photographic study of spatial structure. In A. F. M. Moorwood, editor, *HII regions and the Galactic Centre*, pages 1–11, Proc. 8th ESLAB Symposium, 1974. ESRO SP-105.

[55] J. A. Hackwell, G. L. Grasdalen, and R. D. Gehrz. 10 and 20 micron images of regions of star formation. *Ap. J.*, 252:250–268, 1982.

[56] J. A. Hackwell, G. L. Grasdalen, and R. D. Gehrz. The University of Wyoming integrating preamplifier for InSb detectors. *Bull. Am. Ast. Soc.*, 15:642, 1983.

[57] D. N. B. Hall, R. S. Aikens, R. Joyce, and T. W. McCurnin. Johnson noise limited operation of photovoltaic InSb detectors. *Appl. Opt.*, 14:450–453, 1975.

[58] P. Hall and C. D. Mackay. Faint galaxy number-magnitude counts at high galactic latitude—I. *Mon. Not. R. astr. Soc.*, 210:979–992, 1984.

[59] M. S. Hanner, A. T. Tokunaga, G. J. Veeder, and M. F. A'Hearn. Infrared photometry of the dust in comets. *Astron. J.*, 89:162–169, 1984.

[60] T. Hasegawa. The Orion star forming complex. In M. Peimbert and J. Jugaku, editors, *Star Forming Regions*, pages 123–137, D. Reidel, 1987. IAU Symposium vol. 115.

[61] T. Hasegawa. Star formation associated with high-velocity mass outflows. *Astrophys. Sp. Sci.*, 118:421–434, 1986.

[62] M. Hayashi, T. Hasegawa, I. Gatley, R. Garden, and N. Kaifu. The molecular hydrogen emission associated with the Orion bright bar. *Mon. Not. R. astr. Soc.*, 215:31P–36P, 1985.

[63] A. R. Henden and R. H. Kaitchuk. *Astronomical Photometry.* Van Nostrand Reinhold, 1982.

[64] G. H. Herbig. Stars of low to intermediate mass in the Orion Nebula. In *Symposium on the Orion Nebula to Honor Henry Draper*, pages 64–78, Ann. NY Acad. Sci. vol. 395, 1982.

[65] G. H. Herbig and D. M. Terndrup. The Trapezium cluster of the Orion nebula. *Ap. J.*, 307:609–618, 1986.

[66] A. W. Hoffman. Calibration of self-integrating detectors. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 29–35, Institute for Astronomy, University of Hawaii, 1987.

[67] J. R. Houck. Infrared detectors for space applications. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 108–115, Institute for Astronomy, University of Hawaii, 1987.

[68] J. H. Hough, D. J. Axon, M. G. Burton, I. Gatley, S. Sato, J. Bailey, M. J. McCaughrean, I. S. McLean, T. Nagata, D. Allen, R. P. Garden, T. Hasegawa, M. Hayashi, N. Kaifu, M. Morimoto, and D. Walther. Infrared polarization in OMC-1—discovery of a molecular hydrogen reflection nebula. *Mon. Not. R. astr. Soc.*, 222:629–644, 1986.

[69] A. R. Hyland, D. A. Allen, P. J. Barnes, and M. J. Ward. The distribution and nature of the 2 μm radiation of the inner Orion Nebula. *Mon. Not. R. astr. Soc.*, 206:465–474, 1984. (HABW).

[70] J. A. Jamieson. Infrared technology: advances 1975–84, challenges 1985–94. *Opt. Eng.*, 25:688–697, 1986.

[71] J. R. Janesick, T. Elliott, S. Collins, M. M. Blouke, and J. Freeman. Scientific charge-coupled devices. *Opt. Eng.*, 26:692–714, 1987.

[72] K. J. Johnston, P. Palmer, T. L. Wilson, and J. H. Bieging. The distribution of 6 cm H_2CO in Orion Molecular Cloud 1. *Ap. J.*, 271:L89–L93, 1983.

[73] R. D. Joseph, G. S. Wright, and R. Wade. Detection of molecular hydrogen in two merging galaxies. *Nature*, 311:132–133, 1984.

[74] R. R. Joyce. Indium antimonide detectors for ground-based astronomy. In W. L. Wolfe, editor, *Infrared Detectors*, pages 50–58, Proc. SPIE vol. 443, 1983.

[75] J. Keene, R. H. Hildebrand, and S. E. Whitcomb. A high resolution submillimeter map of OMC-1. *Ap. J.*, 252:L11–L15, 1982.

[76] M. F. Kessler. The status of the ISO project. In C. R. McCreight, editor, *Second Infrared Detector Workshop*, pages 14–1–14–5, NASA Ames Research Center, 1985. NASA (88213).

[77] R. J. Keyes, editor. *Optical and infrared detectors.* Volume 19 of *Topics in Applied Physics*, Springer-Verlag, 1977.

[78] M. Kimata, M. Denda, N. Yutani, S. Iwade, N. Tsubouchi, M. Daido, H. Furukawa, R. Tsunoda, and T. Kanno. 256×256 element platinum silicide Schottky-barrier infrared charge-coupled device image sensor. *Opt. Eng.*, 26:209–215, 1987.

[79] R. H. Kingston. *Detection of optical and infrared radiation.* Volume 10 of *Springer Series in Optical Sciences*, Springer-Verlag, 1978.

[80] W. F. Koso nocky and H. Elabd. Schottky-barrier infrared charge-coupled device focal plane arrays. In W. L. Wolfe, editor, *Infrared Detectors*, pages 167–188, Proc. SPIE vol. 443, 1983.

[81] V. I. Krassovsky, B. P. Potapov, A. I. Semenov, V. G. Sobolev, M. V. Shagaev, and N. N. Shefov. On the equilibrium nature of the rotational temperature of the hydroxyl airglow. *Planet. Space Sci.*, 25:596–597, 1977.

[82] V. I. Krassovsky, N. N. Shefov, and V. I. Yarin. Atlas of the airglow spectrum 3000–12400 Å. *Planet. Space Sci.*, 9:883–915, 1962.

[83] K. Krisciunas, W. Sinton, D. Tholen, A. Tokunaga W. Golisch, D. Griep, C. Kaminski, C. Impey, and C. Christian. Atmospheric extinction and night-sky brightness at Mauna Kea. *Pub. A. S. P.*, 99:887–894, 1987.

[84] F. Lacombe, P. Léna, and D. Rouan. Sub-arcsec imaging of the Galactic Centre in the near infrared. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 316–320, Institute for Astronomy, University of Hawaii, 1987.

[85] P. Laques and J. L. Vidal. Détection de condensations d'un type nouveau dans la centre de la nébuleuse d'Orion. *Astron. Astrophys.*, 73:97–106, 1979.

[86] P. Léna. Observational techniques in infrared astronomy. In G. Setti and G. G. Fazio, editors, *Infrared Astronomy*, pages 231–269, D. Reidel, 1978. NATO Advanced Study Institutes Series vol. 38.

[87] S. J. Lilly and L. L. Cowie. Deep infrared surveys. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 473–482, Institute for Astronomy, University of Hawaii, 1987.

[88] E. J. Llewellyn, B. H. Long, and B. H. Solheim. The quenching of OH* in the atmosphere. *Planet. Space Sci.*, 26:525–531, 1978.

[89] E. J. Llewellyn and B. H. Solheim. The excitation of the infrared atmospheric oxygen bands in the nightglow. *Planet. Space Sci.*, 26:533–538, 1978.

[90] C. J. Lonsdale, E. E. Becklin, T. J. Lee, and J. M. Stewart. New members of the infrared cluster in the Orion molecular cloud. *Astron. J.*, 87:1819–1827, 1982. (LBLS).

[91] J. L. Lorre and A. R. Gillespie. Artifacts in digital images. In D. A. Elliott, editor, *Applications of Digital Image Processing to Astronomy*, pages 123–135, Proc. SPIE vol. 264, 1980.

[92] C. D. Mackay. Charge-coupled devices in astronomy. *Ann. Revs. Astron. Astrophys.*, 24:255–283, 1986.

[93] R. Mauersberger, C. Henkel, T. L. Wilson, and C. M. Walmsley. Hot ammonia in the Galaxy. *Astron. Astrophys.*, 162:199–210, 1986.

[94] M. J. McCaughrean. *Report on simulations and measurements of the emissivity of the United Kindom Infrared Telescope.* Royal Observatory, Edinburgh, June 1987.

[95] M. J. McCaughrean and I. S. McLean. Simulations and modelling of infrared camera systems. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 193–196, Institute for Astronomy, University of Hawaii, 1987.

[96] R. A. McClatchey, R. W. Fenn, J. E. A. Selby, F. E. Volz, and J. S. Garing. *Optical properties of the atmosphere (third edition).* Environmental research paper 411, Air Force Cambridge Research Laboratories, 1972. AFCRL–72–0497.

[97] C. R. McCreight and J. H. Goebel. Integrated infrared detector arrays for low-background applications. In D. L. Crawford, editor, *Instrumentation in Astronomy IV*, pages 9–17, Proc. SPIE vol. 331, 1982.

[98] I. S. McLean. Results with the UKIRT infrared camera. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 180–192, Institute for Astronomy, University of Hawaii, 1987.

[99] I. S. McLean, T. C. Chuter, M. J. McCaughrean, and J. T. Rayner. System design of a 1–5 μm camera for astronomy. In D. L. Crawford, editor, *Instrumentation in Astronomy VI*, pages 430–437, Proc. SPIE vol. 627, 1986.

[100] J. Meaburn. An extended, high speed, ionized flow in the vicinity of a Herbig-Haro-like knot (M42-HH2) in the core of the Orion nebula (NGC1976). *Astron. Astrophys.*, 164:358–363, 1986.

[101] A. B. Meinel. OH emission bands in the spectrum of the night sky I. *Ap. J.*, 111:555–564, 1950.

[102] A. B. Meinel. OH emission bands in the spectrum of the night sky II. *Ap. J.*, 112:120–130, 1950.

[103] A. M. Milton. Readout mechanisms for infrared focal plane arrays. In W. L. Wolfe, editor, *Infrared Detectors*, pages 112–119, Proc. SPIE vol. 443, 1983.

[104] M. H. Minshull and S. E. Botts. *Linearity of InSb direct integration detectors used in astronomy applications.* Internal Memorandum 2113-020, Santa Barbara Research Center, November 1984.

[105] J. M. Mooney and E. L. Dereniak. Comparison of the performance limit of Schottky-barrier and standard infrared focal plane arrays. *Opt. Eng.*, 26:223–227, 1987.

[106] A. F. M. Moorwood. IRSPEC : design, performance, and first scientific results. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 379–387, Institute for Astronomy, University of Hawaii, 1987.

[107] G. Moreels and M. Herse. Photographic evidence of waves around the 85 km level. *Planet. Space Sci.*, 25:265–273, 1977.

[108] G. Moreels, G. Megie, A. Vallance Jones, and R. L. Gattinger. An oxygen-hydrogen atmosphere model and its application to the OH emission problem. *Jnl. Atmos. Terr. Phys.*, 39:551–570, 1977.

[109] D. Morrison, R. E. Murphy, D. P. Cruikshank, W. M. Sinton, and T. Z. Martin. Evaluation of Mauna Kea, Hawaii, as an observatory site. *Pub. A. S. P.*, 85:255–267, 1973.

[110] C. M. Mountain. *Astronomical spectrometry in the near-infrared.* PhD thesis, University of London, 1983. Imperial College of Science and Technology.

[111] C. M. Mountain, S. K. Leggett, M. J. Selby, D. E. Blackwell, and A. D. Petford. Measurement of the absolute flux from Vega at 4.92 μm. *Astron. Astrophys.*, 151:399–402, 1985.

[112] G. Münch and K. Taylor. On the spectrum of neutral oxygen in the Orion Nebula. *Ap. J.*, 192:L93–L95, 1974.

[113] G. Münch and O. C. Wilson. On the structure of the Orion Nebula. *Z. Astrophys.*, 56:127–137, 1962.

[114] L. G. Mundy, N. Z. Scoville, L. B. Bååth, C. R. Masson, and D. P. Woody. Protostellar condensations within the Orion ridge. *Ap. J.*, 304:L51–L55, 1986.

[115] J. F. Noxon. The near infrared nightglow continuum. *Planet. Space Sci.*, 26:191–192, 1978.

[116] G. Orias, A. W. Hoffman, and M. Casselmann. 58 × 62 InSb focal plane array for infrared astronomy. In *Instrumentation in Astronomy VI*, pages 408–417, Proc. SPIE vol. 627, 1986.

[117] P. P. Parenago. *Astrometry of stars in Orion.* Technical Report 25, Trudy Sternberg Astr. Inst., 1954.

[118] A. J. Penny. Electronographic UBV photometry of close visual double stars. *Mon. Not. R. astr. Soc.*, 187:829–837, 1979.

[119] D. H. Pommerrenig. Extrinsic silicon focal plane arrays. In W. L. Wolfe, editor, *Infrared Detectors*, pages 144–150, Proc. SPIE vol. 443, 1983.

[120] J. T. Rayner. *Aspects of the development of a 1–5 μm camera for astronomy.* PhD thesis, University of Edinburgh, 1988. (in preparation).

[121] M. B. Reine, A. K. Sood, and T. J. Tredwell. *Photovoltaic infrared detectors*, chapter 6. Volume 18 (Mercury Cadmium Telluride) of *Semiconductors and Semimetals*, Academic Press, 1981. R. K. Willardson and A. C. Beer, Eds.

[122] M. J. Rieke, G. H. Rieke, and E. F. Montgomery. Rockwell HgCdTe arrays as imagers. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 213–221, Institute for Astronomy, University of Hawaii, 1987.

[123] F. E. Roach and J. L. Gordon. *The Light of the Night Sky.* Volume 4 of *Geophysics and Astrophysics Monographs*, D. Reidel, 1973.

[124] J. P. Rode, J. D. Blackwell, M. A. Blessinger, and K. Vural. SWIR HgCdTe focal plane arrays for astronomy. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 13–20, Institute for Astronomy, University of Hawaii, 1987.

[125] F. P. Schloerb, P. F. Goldsmith, and N. Z. Scoville. Molecular line mapping of OMC-1. In R. S. Roger and P. E. Dewdney, editors, *Regions of Recent Star Formation*, pages 439–444, D. Reidel, 1982. Astrophysics and Space Science Library vol. 93.

[126] R. Schoolar and E. Tenescu. Analysis of InSb photodiode low temperature characteristics. In H. Nakamura, editor, *Infrared Detectors, Sensors, and Focal Plane Arrays*, pages 2–11, Proc. SPIE vol. 686, 1986.

[127] F. Sibille, C. Cesarsky, and D. Rouan. Review of arrays developed for the Infrared Space Observatory (ISO). In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 116–120, Institute for Astronomy, University of Hawaii, 1987.

[128] F. Sibille, P. Léna, A. Chelli, and D. Stefanovich. Two-dimensional infrared speckle interferometry with a 32 × 32 InSb charge-injection device (CID) array. In D. L. Crawford, editor, *Instrumentation in Astronomy IV*, pages 26–28, Proc. SPIE vol. 331, 1982.

[129] J. Smith, A. Bentley, M. Castelaz, R. D. Gehrz, G. L. Grasdalen, and J. A. Hackwell. Infrared sources and excitation of the W40 complex. *Ap. J.*, 291:571–580, 1985.

[130] V. G. Sobolev. Continuum in night airglow between 8000 and 11000 Å. *Planet. Space Sci.*, 26:703–704, 1978.

[131] B. T. Soifer and J. L. Pipher. Instrumentation for infrared astronomy. *Ann. Revs. Astron. Astrophys.*, 16:335–369, 1978.

[132] A. T. Stair, Jr., E. R. Huppi, B. P. Sandford, R. E. Murphy, R. R. O'Neil, A. M. Hart, R. J. Huppi, and W. R. Pendleton. Infrared investigations of the aurora and airglow. In B. M. McCormac, editor, *The Radiating Atmosphere*, pages 185–204, D. Reidel, 1971. Astrophysics and Space Science Library vol. 24.

[133] J. M. Stewart. *Near infrared surface photometry of spiral galaxies.* PhD thesis, University of Edinbugh, 1984.

[134] S. M. Sze. *Physics of Semiconductor Devices.* Wiley-Interscience, 1981.

[135] H. Takahashi and P. P. Batista. Comments on the paper entitled "The quenching of OH* in the atmosphere" by Llewellyn, Long, andSolheim. *Planet. Space Sci.*, 28:559–561, 1980.

[136] H. Takahashi, P. P. Batista, Y. Sahai, and B. R. Clemesha. Atmospheric wave propagations in the mesopause region observed by the OH(8,3) band, NaD, O_2A(8645 Å) band and OI 5577 Å nightglow emissions. *Planet. Space Sci.*, 33:381–384, 1985.

[137] H. Takahashi, Y. Sahai, B. R. Clemesha, P.P. Batista, and N. R. Teixeira. Diurnal and seasonal variations of the OH(8,3) airglow band and its correlation with OI 5577Å. *Planet. Space Sci.*, 25:541–547, 1977.

[138] K. Taylor, J. E. Dyson, D. J. Axon, and S. Hughes. The Herbig-Haro-like objects in M42. *Mon. Not. R. astr. Soc.*, 221:155–168, 1986.

[139] K. Taylor and G. Münch. High velocity features in M42 driven by stellar winds. *Astron. Astrophys.*, 70:359–366, 1978.

[140] M. J. Taylor, M. A. Hapgood, and P. Rothwell. Observations of gravity wave propagation in OI(557.7nm), Na(589.2nm) and the near infrared OH nightglow emissions. *Planet. Space Sci.*, 35:413–427, 1987.

[141] Infrared Array Project Team. *Technical Design Summary for IRCAM.* Royal Observatory, Edinburgh, 1986.

[142] R. D. Thom and B. T. Yang. Low doped InSb arrays for space astronomy. In C. R. McCreight, editor, *Second Infrared Detector Workshop*, pages 10–1–10–12, NASA Ames Research Center, 1985. NASA (88213).

[143] R. I. Thompson, M. J. Lebofsky, and G. H. Rieke. The 2–2.5 micron spectrum of NGC1068 : A detection of extragalactic molecular hydrogen. *Ap. J.*, 222:L49–L53, 1978.

[144] A. T. Tokunaga. *IRTF Photometry Manual.* Infrared Telescope Facility, 1986.

[145] W. A. Traub and M. T. Stier. Theoretical atmospheric transmission in the mid- and far-infrared at four altitudes. *Appl. Optics*, 15:364–377, 1976.

[146] J. A. Tyson. Low-light-level charge-coupled device imaging in astronomy. *J. Opt. Soc. Am. A*, 3:2131–2138, 1986.

[147] A. Vallance Jones. *The IR spectrum of the nightglow*, pages 355–400. Volume 15 of *Space Science Reviews*, D. Reidel, 1973.

[148] A. Vallance Jones and R. L. Gattinger. IR spectrum of aurora. In B. M. McCormac, editor, *The Radiating Atmosphere*, pages 176–184, D. Reidel, 1971. Astrophysics and Space Science Library vol. 24.

[149] J. L. Vampola. *InSb detector modelling for ADS.* Internal Memorandum 2118–244, Santa Barbara Research Center, September 1982.

[150] R. Wade. A 1–5 μm cooled grating array spectrometer and Fabry-Pérot system for UKIRT. In A. Boksenberg and D. L. Crawford, editors, *Instrumentation in Astronomy V*, pages 47–50, Proc. SPIE vol. 445, 19.

[151] R. Wade, T. J. Lee, I. S. McLean, C. M. Mountain, M. J. McCaughrean, and I. Baker. Desired characteristics of 2-D arrays for SWIR imagers and spectrographs. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 517–521, Institute for Astronomy, University of Hawaii, 1987.

[152] M. F. Walker. A comparison of observing conditions on the summit cones and shield of Mauna Kea. *Pub. A. S. P.*, 95:903–918, 1983.

[153] J. R. Walsh. Complex ionized structure in the θ^2 Orionis region. *Mon. Not. R. astr. Soc.*, 201:561–567, 1982.

[154] M. W. Werner. Infrared studies of star formation in Orion. In *Symposium on the Orion Nebula to Honor Henry Draper*, pages 79–99, Ann. NY Acad. Sci. vol. 395, 1982.

[155] M. W. Werner. SIRTF — the Space Infrared Telescope Facility. In C. R. McCreight, editor, *Second Infrared Detector Workshop*, pages 13–1–13–7, NASA Ames Research Center, 1985. NASA (88213).

[156] M. W. Werner and E. F. Erickson. Future space and sub-orbital platforms that will complement ground-based astronomy with infrared arrays. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 527–535, Institute for Astronomy, University of Hawaii, 1987.

[157] R. H. Wiens and G. Weill. Diurnal, annual, and solar cycle variations of hydroxyl and sodium nightglow intensities in the Europe-Africa sector. *Planet. Space Sci.*, 21:1011–1027, 1973.

[158] R. K. Willardson and A. C. Beer, editors. *Infrared detectors.* Volume 5 of *Semiconductors and Semimetals*, Academic Press, 1970.

[159] R. K. Willardson and A. C. Beer, editors. *Infrared detectors II.* Volume 12 of *Semiconductors and Semimetals*, Academic Press, 1977.

[160] T. L. Wilson and T. Pauls. Radio continuum and recombination line observations of Orion A. *Astron. Astrophys.*, 138:225–230, 1984.

[161] W. L. Wolfe and G. J. Zissis, editors. *The IR Handbook.* Office of Naval Research, Department of the Navy, 1978.

[162] C. E. Woodward, A. Moneti, J. L. Pipher, and W. J. Forrest. Arc-second imaging of the BN-KL star formation region. In J. P. Hearnshaw and P. L. Cottrell, editors, *Instrumentation and Research Programmes for Small Telescopes*, pages 325–326, IAU Symposium vol. 118, D. Reidel, 1986.

[163] P. C. Wraight. The near infrared nightglow continuum. *Planet. Space. Sci.*, 25:787–794, 1977.

[164] J. F. Wright and C. D. Mackay. The Cambridge charge-coupled device (CCD) system. In J. C. Geary and D. W. Latham, editors, *Solid State Imagers for Astronomy*, pages 160–164, Proc. SPIE vol. 290, 1981.

[165] C. G. Wynn-Williams and E. E. Becklin, editors. *Infrared Astronomy with Arrays*, Institute for Astronomy, University of Hawaii, 1987.

[166] C. G. Wynn-Williams, R. Genzel, E. E. Becklin, and D. Downes. The Kleinmann-Low nebula : an infrared cavity. *Ap. J.*, 281:172–183, 1984.

[167] E. S. Yang. *Fundamentals of Semiconductor Devices.* McGraw-Hill, 1978.

[168] A. T. Young. Seeing : its cause and cure. *Ap. J.*, 189:567–604, 1974.

[169] L. M. Ziurys and P. Friberg. Observations of SiO towards OMC-1 : a new outflow source 1.5$\hat{}$ south of Orion-KL ? *Ap. J.*, 314:L49–L53, 1987.

Appendix A

Source positions

In this appendix we give astrometric positions for many of the sources discussed in chapter 5. The positions were derived from far red Schmidt plates and the IRCAM K mosaic of the Orion Nebula as discussed in that chapter. The positions are given at both B1950.0 and J2000.0 equinoxes. The absolute accuracy of the positions is $\pm 1''$, whilst the relative accuracy within the K mosaic is $\pm 0.6''$.

Positions for two of the bright optical stars θ^1 Ori C and θ^2 Ori A are given first. Then we list the positions of 22 of the 26 LBLS sources [90] which are seen to be pointlike or only slightly extended, including BN. The 4 infrared sources close to the line of sight towards the southern molecular condensation OMC-1S are then listed, named S1 to S4, as in the text. The nebulosities around θ^2 Ori A are listed by the names given by Walsh [153], whilst S5 is the the infrared point source located between knots A and B. Sources S6 to S9 are associated with the Herbig-Haro source M42-HH2, as discussed in the text, and S10 and S11 are associated with M42-HH1.

Name	B1950.0		J2000.0		Other names
	RA	Dec	RA	Dec	
Bright stars					
θ^1 Ori C	05 32 49.02	−05 25 16.7	05 35 16.48	−05 23 23.4	
θ^2 Ori A	05 32 55.44	−05 26 51.8	05 35 22.87	−05 24 59.0	
LBLS sources					
BN	05 32 46.63	−05 24 17.3	05 35 14.12	−05 22 23.9	IRc1
a	05 32 45.51	−05 24 09.8	05 35 13.00	−05 22 16.2	
b	05 32 45.67	−05 24 16.0	05 35 13.16	−05 22 22.5	
c	05 32 45.84	−05 24 20.4	05 35 13.32	−05 22 26.9	
d	05 32 46.04	−05 24 31.6	05 35 13.52	−05 22 38.1	
e	05 32 46.01	−05 24 14.2	05 35 13.49	−05 22 20.7	π1820
g	05 32 46.26	−05 24 16.0	05 35 13.74	−05 22 22.6	π1821
h	05 32 46.27	−05 24 01.7	05 35 13.76	−05 22 08.2	π1819
i	05 32 46.42	−05 24 24.1	05 35 13.90	−05 22 30.7	IRc3
k	05 32 46.62	−05 24 30.4	05 35 14.10	−05 22 36.9	π1822
m	05 32 46.77	−05 23 59.3	05 35 14.26	−05 22 05.8	
n	05 32 46.88	−05 24 27.3	05 35 14.36	−05 22 33.9	IRc4(?)
p	05 32 46.88	−05 24 30.4	05 35 14.35	−05 22 37.0	
r	05 32 46.96	−05 24 23.6	05 35 14.45	−05 22 30.1	IRc2
s	05 32 46.98	−05 24 01.1	05 35 14.47	−05 22 07.7	
t	05 32 47.17	−05 24 27.9	05 35 14.65	−05 22 34.5	π1839
u	05 32 47.38	−05 24 25.5	05 35 14.86	−05 22 32.1	
v	05 32 47.46	−05 24 33.6	05 35 14.94	−05 22 40.2	π1840
w	05 32 47.72	−05 24 24.2	05 35 15.20	−05 22 30.9	
x	05 32 47.67	−05 24 30.5	05 35 15.15	−05 22 37.1	
y	05 32 47.85	−05 24 19.3	05 35 15.33	−05 22 25.9	π1838
z	05 32 47.89	−05 24 19.3	05 35 15.37	−05 22 25.9	
OMC-1S sources					
S1	05 32 46.15	−05 25 49.5	05 35 13.60	−05 23 56.0	
S2	05 32 46.10	−05 25 53.8	05 35 13.55	−05 24 00.3	
S3	05 32 45.52	−05 25 47.5	05 35 12.97	−05 23 54.0	
S4	05 32 46.33	−05 25 33.9	05 35 13.78	−05 23 40.4	
Nebulosity near θ^2 Ori A					
Knot A	05 32 54.76	−05 26 54.9	05 35 22.19	−05 25 02.0	M42-HH3
Knot B	05 32 55.25	−05 27 12.4	05 35 22.68	−05 25 19.5	M42-HH4
Knot C	05 32 57.04	−05 26 33.2	05 35 24.48	−05 24 40.5	π2011
Knot E	05 32 58.40	−05 26 08.4	05 35 25.84	−05 24 15.8	
Knot F	05 32 54.08	−05 27 09.2	05 35 21.51	−05 25 16.2	
S5	05 32 54.97	−05 27 03.6	05 35 22.40	−05 25 10.8	
M42-HH2 sources					
S6	05 32 44.52	−05 24 48.9	05 35 11.99	−05 22 55.3	
S7	05 32 44.14	−05 24 45.8	05 35 11.62	−05 22 52.1	
S8	05 32 44.26	−05 24 51.4	05 35 11.74	−05 22 57.8	
S9	05 32 45.10	−05 24 55.8	05 35 12.57	−05 23 02.3	
M42-HH1 sources					
S10	05 32 44.27	−05 23 51.6	05 35 11.76	−05 21 58.0	
S11	05 32 43.94	−05 23 49.1	05 35 11.43	−05 21 55.4	

Figure A.1: Astrometric positions for sources in IRCAM K mosaic of Orion

Appendix B

The Rapi2d software package

As part of my work for the Infrared Array Project, I have been responsible for the implementation of a software package designed to handle the basic data reduction and analysis requirements for data from the infrared camera, IRCAM. The package is known as Rapi2d, for Reduction and Analysis of Infrared Imaging Data.

This package was implemented in the ADAM environment, which is a multi-task system designed to interface telescope, instrument, data storage, reduction, and display tasks at the United Kingdom's telescopes in Hawaii, namely the United Kingdom Infrared Telescope, and the James Clerk Maxwell Telescope. The system has since been adopted for control of the William Herschel Telescope in the Canaries, the Anglo-Australian Telescope in Australia, and has also been adopted by Starlink, to provide a full off-line image and data processing environment in the United Kingdom.

Rapi2d contains all the basic features required of an on-line image processing system, with the graphics provided by Colin Aspin's PLT2D package. These features include basic image arithmetic, smoothing, shape modification, bad pixel removal, image mosaicing, thresholding, statistics, basic feature enhancement, and manipulation of image orientation. It does not attempt to duplicate the more difficult tasks of off-line data analysis, which include interactive and batch aperture photometry, galaxy photometry, profile fitting, image deconvolution, and so on. These routines already exist in other packages, such as Figaro, Aspic, AIPS, DAOPHOT, and so on. Starlink are responsible for collecting together these other packages as part of their general purpose image processing environment, and for providing new packages required under the ADAM

environment.

The data format used by Rapi2d is the Starlink Data Format (or SDF), which uses the Hierarchical Data System (HDS) to provide a flexible method of storing bulk data, such as images. The system allows for an almost infinite variety of data formats, each of which can be tailored to suit a particular application. For Rapi2d however, a basic format has been adopted that consists simply of the data and the image dimensions. In this way, Rapi2d should be able to process any two-dimensional data in SDF format, as it does not presume the existence of any special components.

It is worth noting that image simulation routines in the SIRCAM package are also implemented in the ADAM environment, and that the simultaneous development of Rapi2d and SIRCAM allowed simulated data to be analysed, whilst providing feedback to the development of the reduction package.

I acknowledge the work Dave Baines of the ROE, and Keith Hartley of RGO, whose work for the now defunct Starlink Software Environment was used as a template for Rapi2d. Finally, it should be noted that Rapi2d has itself since been used as a template for KAPPA, the new Starlink kernel package for general purpose image processing.

A list of the Rapi2d routines and a one-line summary of the purpose of each is given below.

List of routines

ADD	:	add image+image into new output image
APERADD	:	bins up pixels in a circular aperture and gives statistics
BLOCK	:	perform a block smooth on a 2-D image
CADD	:	add image+scalar into new output image
CDIV	:	divide image/scalar into new output image
CENTROID	:	locates star centres by centroiding technique
CHPIX	:	replace 'bad' pixels with specified values
CMULT	:	multiply image×scalar into new output image
COMPADD	:	compresses image by adding together a specified number of pixels
COMPAVE	:	compresses image by averaging over specified number of pixels

COMPICK	:	compresses image by selecting pixels from the input image
COMPRESS	:	compresses image by different amounts in x and y dimensions
CREFRAME	:	allows user to generate specific forms of test data
CSUB	:	subtract image-scalar into new output image
DIV	:	divide image/image into new output image
EXP10	:	take exponential of of image into new output image (base 10)
EXPE	:	take exponential of of image into new output image (base e)
EXPON	:	take exponential of of image into new output image (specified base)
FLIP	:	reverse an image about either the horizontal or vertical axis
GAUSS	:	perform a Gaussian smooth on a 2-D image
GLITCH	:	replace 'bad' pixels with local median
HISTEQ	:	performs histogram equalisation on an image
HISTO	:	calculates histogram of an image sub-array and statistical parameters
LAPLACE	:	perform a Laplacian convolution as an edge detector
LOGAR	:	take logarithm of image into new output image (specified base)
LOGE	:	take logarithm of image into new output image (base e)
LOGTEN	:	take logarithm of image into new output image (base 10)
LOOK	:	allows inspection of image values to screen or listing file
MANIC	:	1,2 or 3-D image to 1,2 or 3-D image reshaping program
MASK	:	mask input image by defined mask into output image
MEDIAN	:	perform weighted median filtering on a 2-D image
MOSAIC	:	merges several non-congruent images into one big image
MSTATS	:	works out image statistics for a sequence of frames
MULT	:	multiply image×image into new output image
NUMB	:	counts number of pixels with absolute values $\geq$ value
OUTSET	:	replaces pixels outside specified circle with specified value
PICKIM	:	creates a new image from a subset of another image
PIXDUPE	:	expands an input image by pixel duplication
POWER	:	take the power of image into new output image (specified power)
QUILT	:	sophisticated mosaic routine
ROTATE	:	rotate an image through any number of degrees
SHADOW	:	enhances an image by simulating side-lighting effect
SHIFT	:	realign an image

SHSIZE	:	output the dimensions of an image to the environment
SQORST	:	change the dimensions of an image
STATS	:	works out image statistics
SUB	:	subtract image-image into new output image
THRESH	:	threshold all values outside a given range to user defined values
THRESH0	:	threshold all values outside a given range to zero
TRACE	:	trace through the contents of a data structure
TRANDAT	:	converts free-format x,y,intensity data into .SDF images
TRIG	:	performs trigonometrical transformations on images
ZAPLIN	:	interpolates over bad rows or columns

Appendix C

Published papers

This appendix is comprised of copies of all papers involving the author, published during his postgraduate studies. The papers are :

- I. S. McLean, C. Aspin, S. R. Heathcote, and M. J. McCaughrean. Is the polarization of NGC1068 evidence for a non-thermal source? *Nature*, 304:609–611, 1983.

- C. Aspin, I. S. McLean, and M. J. McCaughrean. CCD observations of bipolar nebulae II. Lk Hα233. *Astron. Astrophys.*, 144:220–226, 1985.

- J. H. Hough, D. J. Axon, M. G. Burton, I. Gatley, S. Sato, J. Bailey, M. J. McCaughrean, I. S. McLean, T. Nagata, D. Allen, R. P. Garden, T. Hasegawa, M. Hayashi, N. Kaifu, M. Morimoto, and D. Walther. Infrared polarization in OMC-1—discovery of a molecular hydrogen reflection nebula. *Mon. Not. R. astr. Soc.*, 222:629–644, 1986.

- I. S. McLean, T. C. Chuter, M. J. McCaughrean, and J. T. Rayner. System design of a 1–5 μm IR camera for astronomy. In D. L. Crawford, editor, *Instrumentation in Astronomy VI*, pages 430–437, Proc. SPIE vol. 627, 1986.

- I. S. McLean, M. J. McCaughrean, I. Gatley, J. Hough, S. Sato, T. Nagata, D. Axon, M. G. Burton, R. Garden, C. Aspin, T. Hasegawa, M. Hayashi, M. Morimoto, and N. Kaifu. Discovery of an extended infrared reflection nebula around S106. *Mon. Not. R. astr. Soc.*, 225:393–398, 1987.

- M. Tamura, T. Nagata, S. Sato, M. Tanaka, N. Kaifu, J. Hough, I. McLean, I. Gatley, R. Garden, and M. McCaughrean. Magnetic field structure in dark clouds.

In M. Peimbert and J. Jugaku, editors, *Star Forming Regions*, pages 48–50, IAU Symposium vol. 115, D. Reidel, 1987.

- M. J. McCaughrean and I. S. McLean. Simulation and modelling of infrared camera systems. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 193–196, Institute for Astronomy, University of Hawaii, 1987.

- R. Wade, T. J. Lee, I. S. McLean, C. M. Mountain, M. J. McCaughrean, and I. Baker. Desired characteristics of 2-D arrays for SWIR imagers and spectrographs. In C. G. Wynn-Williams and E. E. Becklin, editors, *Infrared Astronomy with Arrays*, pages 517–521, Institute for Astronomy, University of Hawaii, 1987.

DAMAGED TEXT IN ORIGINAL

from IRc2. This hypothesis conflicts with the usual interpretation by most authors (e.g. Dyck & Beichman 1974) that the polarization of BN is due to grain alignment via paramagnetic relaxation.

The only near-infrared spectropolarimetric observations have been presented by Joyce & Simon (1982) who measured the linear polarization of the $v=1$–$0\,S(1)$ line of molecular hydrogen at a few points. At the region of peak emission of the line (Peak 1 of BNPB's map) they found the line to be around 10 per cent polarized, in good agreement with the *K*-band polarization (Lonsdale *et al.* 1980). We note, however, that they found a difference in position angle of 30° between the continuum and line polarization at Peak 1. To the SE of BN (Peak 2) the emission-line polarization was reported to be close to zero in contrast to the high continuum levels measured in that region by WDC.

In this paper we present and discuss an extensive new set of broad band and molecular-hydrogen-line polarization measurements made at both high spatial resolution over the central 20 arcsec and at lower spatial resolution out to 80 arcsec from BN. These new observations provide a significant new insight into the physical processes occurring in the Orion Molecular Cloud and reveal evidence for aligned grains and for scattered molecular hydrogen emission. In addition they may have important implications for the interpretation of the kinematics of the high-velocity flow.

2 Observations

Two separate telescopes and polarimeter systems were used to obtain the observations reported here.

High spatial resolution polarization maps at *K* (2.2 μm) were made of the core region of Orion-KL on 1982 December 31 and 1983 January 1 using the 3.9-m Anglo–Australian Telescope (AAT) and the Hatfield Polarimeter (Bailey & Hough 1982) in single-beam mode. A half-waveplate, achromatic between 1.0 and 2.4 μm, continuously rotating at 1.25 Hz, is followed by a dichroic mirror which reflects the beam into the AAT Infrared Photometer Spectrometer (IRPS) within which was mounted a wire-grid analyser. Instrumental polarization was measured at the 0.1 per cent level for the entire system and is therefore completely negligible for the observations reported here. The calibration of the polarization levels and the zero-point of position angles were determined from observations of BN itself.

The data were recorded using a computer-controlled polarimetric mapping system in which the telescope was stepped automatically through a grid of points with data from three rotations of the waveplate being recorded at each position. A complete scan of 400 to 500 spatial positions took about 20 min and four or five scans were combined to produce each map. Each map used a 4.0 arcsec aperture and a grid of points with 2.5 arcsec spacing was scanned. The *K*-map presented here was obtained on 1983 January 1. The map was obtained without spatial chopping and the background levels were derived from the faintest positions on the map.

Lower spatial resolution maps at *K* of the fainter outer regions of the Orion-KL nebula and polarization maps in the *S*(1) line at 2.122 μm were made using the 3.8-m United Kingdom Infrared Telescope (UKIRT) on Mauna Kea, Hawaii, between 1984 November 15–17 and on 1984 December 8. For these observations the Kyoto/UKIRT infrared polarimeter was used with the InSb photometer UKT9. The polarimeter consists of a continuously rotating zero-order half-waveplate (for the *K*-band) which is placed in the *f*/36 telescope beam just above the dichroic mirror which feeds the beam to UKT9. A warm wire-grid polarizer was mounted over the entrance window of the photometer. Modulation of the polarization is performed slowly (~0.125 Hz) while the a.c. signal from UKT9 was handled by phase-sensitive detection using a chopper frequency of 3.5 Hz. Calibration tests on an unpolarized star with and without a wire-grid

polarizer above the waveplate showed that the polarization efficiency was 91.4 per cent across the 2.2 μm window and that instrumental polarization for the entire system was less than 0.2 per cent. Position-angle calibration to the equatorial system was made with reference to L1551–IRS5 ($\theta_{2.2}=162°$ was taken).

Advantage was taken of the large beam-size (19.6 arcsec) available with UKT9 to map at 2.2 μm with good signal-to-noise, an extensive region of low surface-brightness emission; about 35 spatial positions were obtained over a 2×3 arcmin field centred on BN. Using the same beam size, 20 positions in this outer region were also measured for polarization in the $v=1$–$0\,S(1)$ line of molecular hydrogen at 2.122 μm. The line was isolated by placing a Fabry–Perot etalon with an effective resolution of ~130 km s^{-1} in front of the wire-grid polarizer and tuning it to the line wavelength. The additional instrumental polarization caused by the introduction of the Fabry–Perot is a few tenths of a per cent, and therefore negligible for these observations. A further 25 positions were measured for *S*(1) polarization in the central part of OMC-1 using a 7.8 arcsec aperture.

Integration times were 4 min per point. Spatial chopping to remove sky background was made 2 arcmin E–W for the brighter points and 4 arcmin E–W for fainter points. Each position was measured in the right and left beams and the results combined to determine the degree and percentage polarization at each point.

For the BN object, continuum photons contribute as much to the total flux when measured

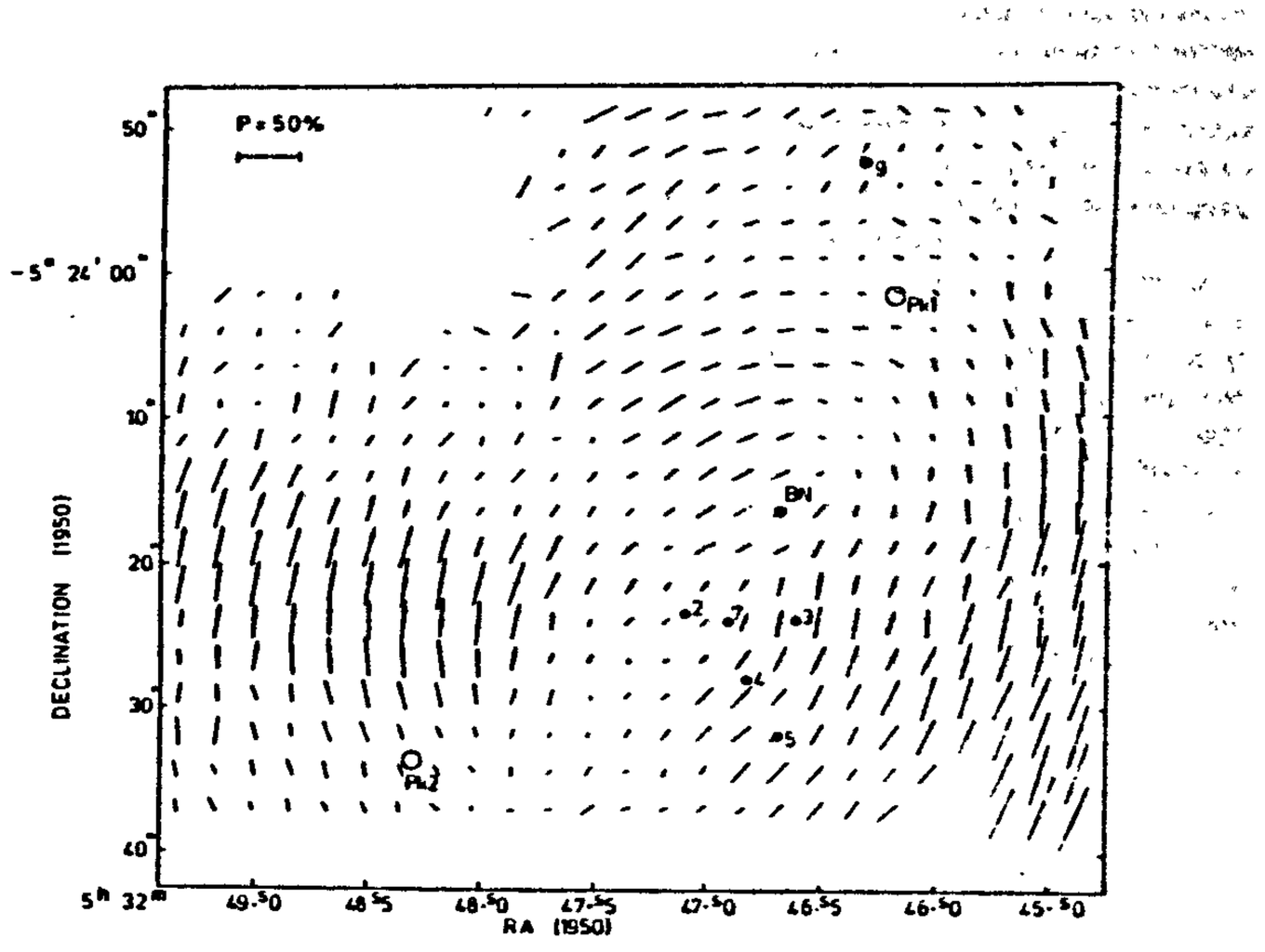


Figure 1. *K*-band polarization map of the BN–KL region in Orion, mapped at the AAT with a 4 arcsec aperture and a grid spacing of 2.5 arcsec. The length of each line is a measure of the polarization and the orientation of the line gives the position angle of the E-vector. Numbers 2, 3, 4, 5, 7 and 9 are IRc numbers (see Wynn-Williams *et al.* 1984 and references therein); BN is the Becklin–Neugebauer object; Pk1 and Pk2 identify the locations of the strongest regions of molecular hydrogen emission in each lobe (Beckwith *et al.* 1978).

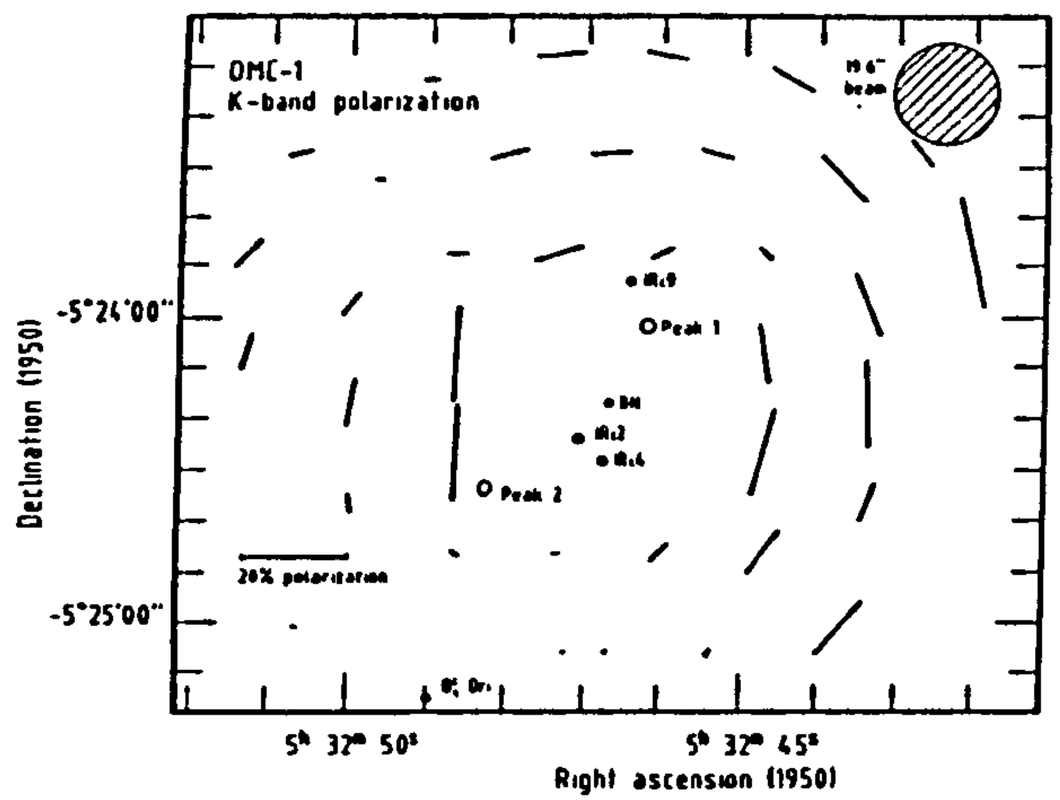


Figure 2. *K*-band polarization map of OMC-1. A 19.6 arcsec aperture was used for this map, taken with the UKIRT. The high-resolution map of Fig. 1 covers the central region of this map. In addition to the sources labelled in Fig. 1, the star θ_1^1 Ori in the Trapezium cluster is shown.

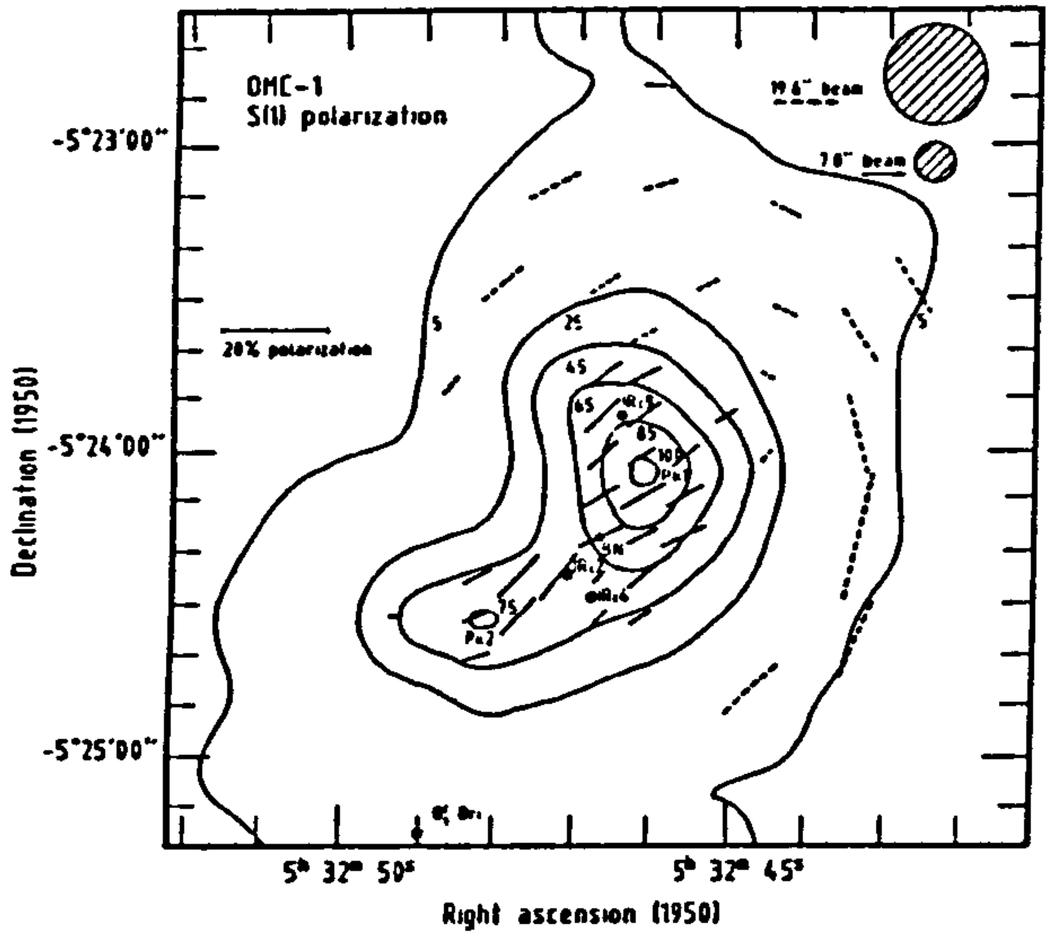


Figure 3. 1–0 *S*(1) line of molecular hydrogen polarization map of OMC-1. Two beam sizes were used for this map obtained with the UKIRT; vectors measured with a 7.8 arcsec beam are shown with a solid line and those measured with a 19.6 arcsec beam with a dashed line. Contour lines are of the *S*(1) intensity as mapped with a 19.6 arcsec beam on the UKIRT. Intensity levels are marked as a percentage of Peak 1's flux. [$3.5 \pm 0.5 \times 10^{-16}$ W cm^{-2} through a 19.6 arcsec beam.]

through the narrow bandpass of the Fabry-Perot as do the *S*(1) line photons. To obtain the line polarization, the continuum polarization was measured with the Fabry-Perot tuned to pass continuum adjacent in wavelength to the *S*(1) line. This was then vector-subtracted from the continuum plus line polarization to yield the resultant line polarization. Elsewhere the effect of continuum photons on the line polarization was negligible.

The results of the AAT and UKIRT observations are shown in three diagrams. Fig. 1 displays the high-resolution map from the AAT. The positions of the infrared sources in the region are indicated on the map. Figs 2 and 3 show the lower resolution larger scale maps of the *K*-band and the *S*(1) line polarization. The *S*(1) map is a composite of measurements made with two different beam sizes, 7.8 and 19.6 arcsec. The *S*(1) intensity contours are taken from the low-resolution molecular hydrogen map of the Orion Molecular Cloud complex by Garden *et al.* (in preparation). Table 1 gives the numerical results for the UKIRT data; offsets in arcsec are referred to the BN object (RA $5^h 32^m 46\overset{s}{.}7$, Dec $-5°24'17''$ 1950.0).

Table 1. Polarization results from the UKIRT.

(a) *K*-band polarization in OMC-1 (through 19.6 arcsec beam).

RA	DEC	P%	ERROR	THETA	ERROR	S/N
-70	30	22.4	0.25	12.1	0.32	90
-60	50	6.4	1.1	40	5	6
-50	-20	7.9	0.5	160	1.8	16
-50	0	17.1	0.3	3	0.4	57
-50	20	13.0	0.7	23	1.5	18
-45	-45	14.4	0.14	141	0.3	103
-45	45	12.2	0.9	45	2.1	13
-35	65	10.0	1.0	60	3	10
-30	-30	10.7	0.28	146	0.8	38
-30	-10	17.3	0.25	165	0.4	69
-30	10	12.2	0.18	9	0.4	68
-30	30	3.7	0.13	46	1	28
-20	-50	1.66	0.01	141	0.3	166
-20	50	7.0	0.3	74	1.4	23
-10	-30	4.0	0.29	137	2	14
-10	30	4.7	0.07	118	0.4	67
-10	70	8.1	0.2	75	0.8	40
0	-50	0.48	0.03	138	2	16
0	50	8.4	0.17	93	0.6	49
8	-50	0.36	0.04	135	3	9
10	-30	1.0	0.10	97	3	10
10	30	9.2	0.16	109	0.5	57
10	70	9.4	0.04	96.4	0.01	235
20	50	7.4	0.5	104	1.9	15
30	-30	1.6	0.05	44	1	32
30	-10	18.6	0.07	177	0.1	265
30	10	18.9	0.08	177	0.1	236
30	30	3.1	0.10	91	1	31
35	65	3.2	0.3	88	3	10
45	45	1.6	0.12	90	2	13
50	-20	3.0	0.02	8	0.2	150
50	0	9.4	0.1	169	0.3	94
50	20	5.9	0.1	141	0.4	59
60	-45	0.66	0.23	66	10	3
60	50	4.6	0.2	102	1	28
70	10	7.3	0.16	164	0.6	45
70	30	7.2	0.3	141	1	24
S(1) Line + Continuum at BN (through 7.8" beam)						
0	0	14.3	1.0	113	2	15
Adjacent Continuum through FP to S(1) Line at BN						
0	0	21.0	2.8	118	4	7

Table 1–*continued*

(b) *S*(1) line polarization in OMC-1.

RA	DEC	BEAM	PT	ERROR	THETA	ERROR	S/N
-24	24	7.8	4 6	1.6	123	10	3
-16	0	7.8	9.4	1.6	117	5	6
-16	8	7.8	7.8	0.8	120	3	10
-16	16	7.8	6.3	0.9	129	4	7
-8	-16	7.8	4.6	1.7	126	11	2
-8	-8	7.8	6.3	2.6	132	12	2
-8	0	7.8	6 9	1.4	120	6	5
-8	8	7.8	8.7	1.0	120	4	8
-8	16	7.8	4.8	0.4	121	3	11
-8	24	7.8	7.5	1.0	129	4	7
-8	32	7.8	5.9	1.5	115	8	4
0	-16	7.8	3.9	1.7	138	12	2
0	-8	7.8	3.2	1.5	135	13	2
0	0	7.8	8.8	1.5	117	5	6
0	8	7.8	7.7	1.2	120	5	6
0	16	7.8	8.5	1.4	135	5	6
0	24	7.8	9.8	1.0	134	3	10
0	32	7.8	8.8	3.4	125	11	2
8	-8	7.8	10.2	2.5	144	7	4
16	-16	7.8	9.1	1.7	139	6	5
16	-8	7.8	10.3	1.5	137	5	7
24	-24	7.8	6.6	2 0	107	9	3
24	-16	7.8	5.1	1.2	112	7	4
24	-8	7.8	6.2	2.0	116	10	3
-60	50	19.6	12.7	3	32	7	4
-50	-20	19.6	15.5	0.2	159	0.3	77
-50	0	19.6	24.9	1	169	1	25
-50	20	19.6	16.7	0.7	14	1.2	24
-50	40	19.6	12.3	0.7	32	1.6	17
-35	45	19.6	3.9	0.5	61	4	8
-35	65	19.6	7.1	1.8	65	7	4
-32	16	19.6	2.9	1.4	140	13	2
-32	32	19.6	2.1	0.8	64	11	2
-30	-30	19.6	14.6	2.4	134	5	6
-20	50	19.6	4.3	0.8	115	5	5
-10	70	19.6	6.0	0.4	105	2	15
-10	90	19.6	6.7	1	86	4	6
-8	40	19.6	5.1	0.8	120	5	6
0	0	19.6	12.3	1	112	2.5	12
0	50	19.6	6.1	0.7	122	3	8
10	70	19.6	10.7	1	119	3	10
20	50	19.6	9.6	1	132	3	9
30	30	19.6	4.7	0.4	139	2.3	12
40	-16	19.6	2.3	1.9	88	20	1

3 Results

3.1 CONTINUUM (BROAD-BAND) POLARIZATION

Although line emission will contaminate the 2 μm data to some extent, the broad-band observations mainly characterize the continuum radiation. At Peak 1 of the molecular hydrogen emission we estimate that the line contributes only 2 per cent of the total flux at K in the 4 arcsec beam of the AAT. The broad-band polarization maps presented in Figs 1 and 2 are considerably more extensive than those of WDC and show some new features. The overall pattern is quite complex, but can be characterized by an approximate symmetry of the vectors around the core region containing IRc2 and BN. The centro-symmetric vectors are typical of reflection, and this

pattern is seen more clearly as we move out from the core region. From Fig. 2 (the UKIRT map) we confirm that the polarization here (mainly) arises by scattering in a reflection nebula centred on the sources IRc2 and BN. The higher resolution AAT data (Fig. 1) highlight IRc2 as the dominant source. The vectors are typically around 8–15 per cent polarized in the outer regions and rise to over 20 per cent in much of the core. These values at K are generally lower than those measured at L' by WDC.

Close to and between BN and IRc2 (see Fig. 1), however, the circular pattern of the E-vectors is disrupted and there is a clear tendency for the polarization vectors to be aligned with position angles similar to those for BN. Apparently some twisting of the position angles occur between BN and IRc2. This result is new and is not consistent with polarization being produced entirely by the scattering of radiation from either source. Moreover, the region to the east and south-east of IRc2 (Fig. 1) is also not consistent with a centro-symmetric scattering pattern. A clear breakdown in the centro-symmetric pattern can also be seen around the self-luminous source IRc9.

These detailed observations imply that there may be more than one polarizing mechanism/geometry. The UKIRT data of Fig. 2 show that low-surface-brightness scattered continuum radiation extends well outside the core region.

3.2 *S*(1) LINE POLARIZATION

Perhaps the most significant result of this paper is the remarkable polarization structure discovered by extensive mapping in the $v=1$–0 $S(1)$ line of molecular hydrogen (H_2) – see Fig. 3. Over the region of high-surface-brightness line emission, i.e. the region covered by the map of BNPB, the polarization in the $S(1)$ line shows a strong alignment of vectors approximately parallel to the *continuum* position angle of BN(118°), with polarization values between 5 and 10 per cent and average position angle about 120°. This direction is parallel to the position angle that the outflow axis of the source makes on the sky, as defined by redshifted and blueshifted CO gas (Erickson *et al.* 1982) and by the two lobes of shocked H_2 emission in BNPB's map [and in our $S(1)$ map]. *(In the rest of the paper we refer to this direction as the outflow direction.)* Beyond this region of strong emission the alignment of polarization vectors ceases fairly abruptly and the centro-symmetric pattern characteristic of scattering emerges. In this outer region the pattern and polarization levels are similar to the K-band but the $S(1)$ polarization vectors are *systematically rotated* with respect to the corresponding continuum vectors. Levels of polarization here are high, exceeding 10 per cent in many places and having a maximum of 25 per cent. A least-squares analysis of the vector orientations for the centro-symmetric vectors of Fig. 3 implies that the source of illumination for the $S(1)$ scattering pattern is a region 18±5 arcsec W, 7±5 arcsec N (standard errors) of BN, i.e. close to the intense H_2 source called Peak 1 on the BNPB map. The implication of this result is that a major fraction of the observed surface brightness of H_2 emission, outside the region of aligned vectors to the north of IRc2/BN, is not intrinsic but is *scattered* into the line-of-sight from an intense relatively compact region.

There is a clear distinction along the outflow axis between the regions of aligned and centro-symmetric polarization. This transition marks the physical boundaries of the regions of high-velocity flow which are intrinsically radiating shocked molecular hydrogen lines. Along the direction perpendicular to the outflow axis it is not easy to distinguish the transition from the aligned pattern to the centro-symmetric pattern due to degeneracy between the vectors. In this region a centro-symmetric vector lies parallel to the outflow axis. However, the data are consistent with all the $S(1)$ emission from the shock-excited region being polarized with position angle parallel to the outflow axis. We note that for the $S(1)$ radiation this alignment of the polarization vectors extends much further out along the direction of the outflow than is observed for the continuum radiation.

In the region between BN and Peak 2 of the BNPB map the pattern of aligned polarization vectors (Fig. 3) is twisted. Near Peak 2 the position angle is ~112°, it rotates to ~144° near IRc2 and then back to ~117° near BN. The maximum distortion of the pattern occurs near IRc2, the likely source of the outflow.

Finally, the continuum-corrected line polarization for BN itself is (9±1.5) per cent, which is significantly lower than the polarization of the continuum either side of the $S(1)$ line, seen to be (21±3) per cent.

4 Discussion

The polarization data presented in this paper contain a wealth of information relating to the sources of illumination in OMC-1, the distribution of gas and dust, the geometry of the outflow and the alignment of grains. By comparing the line and continuum data, the overall impression gained is that the continuum radiation arises from compact sources and is scattered through large angles in a dust halo, whereas the line radiation comes from both compact and extended sources. Radiation from the extended line sources and near the line-of-sight to the continuum and compact line sources evidently reaches us by passing through a medium of aligned dust grains.

There is already considerable evidence in favour of the polarization of BN in the near- and thermal-infrared being due to preferential extinction (dichroic absorption) of one plane of vibration on the passage of radiation through aligned grains. This interpretation is based mainly on the good correlation between the increase in the level of polarization with the increase in extinction across both the silicate and ice-absorption features, the independence of position angle with wavelength, and the high degree of circular polarization (Dyck & Lonsdale 1980). We suggest, therefore, that close to the sources of continuum illumination (IRc2, BN and IRc9), the radiation we observe is scattered through small angles and is thereby only weakly polarized by the scattering. The observed continuum polarization is then mainly produced by an overlying dichroic slab of aligned grains. The tendency for the broad-band polarization to show alignment in the region between IRc2 and BN (Fig. 1) is thus explained. Furthermore, we believe that the strongly aligned polarization pattern observed in the $S(1)$ line radiation (Fig. 3) must similarly result from dichroic absorption. If the line emission is spatially extended then polarization by reflection of $S(1)$ photons is small because any given scatterer 'sees' a more isotropic radiation field. Line emission reaching us is therefore direct or predominantly forward scattered and once again its polarization will be dominated by the dichroic slab. At large angular extent from the line and continuum sources one expects the polarization by reflection to be significant because the scattering angles will be large. Since the observed reflection patterns of vectors are essentially centro-symmetric this suggests that the major sources of line and continuum radiation are localized. Some of these points and their implications are discussed below in more detail.

4.1 Inner (molecular outflow) region

The systematic alignment of the polarization vectors in the outflow region is most obviously explained by dichroic absorption of the $S(1)$ radiation. The pattern is explained naturally if the photons pass through a medium containing aligned, elongated grains. These grains would have to be aligned on average with their long axes perpendicular to the outflow axis. We are observing the radiation through such a slab of material, whose extent must cover at least the molecular outflow region. We discuss the likely extent of the dichroic sheet and possible mechanisms for the grain alignment in later sections.

It could be argued that the pattern in the outflow region is produced by scattering off aligned grains. Scattering of radiation off a medium of aligned grains can produce polarization vectors

lying along the long axis of the grain (Martin 1978). For this mechanism to work, grains must be aligned with their long axes along the outflow axis. The question then arises as to why the observed polarization vectors are aligned in the same direction as for BN. It could not be due to transmission through the same aligned grains (we note that there is strong evidence that the polarization of BN results from dichroic absorption) as then the polarization vector would come out orthogonal to the outflow axis. The coincidence then appears fortuitous and we would be forced to argue for a contrived geometry. We prefer the argument that the alignment of the vectors is due to dichroic absorption of the radiation.

The $S(1)$ and 2.12 μm continuum band position angles at BN are the same (117°), as required for dichroic absorption. However, the percentage polarization for the continuum photons at 2.12 μm is considerably greater than the $S(1)$ line photons at that wavelength (21 per cent compared to 9 per cent), whereas the measured intensities through the 130 km s^{-1} bandpass of the Fabry-Perot are almost the same. For the percentage polarization to be so much larger, the continuum radiation from BN must traverse a greater path length through aligned grains, or a large fraction of the $S(1)$ emission is produced foreground of the aligned grains. If we *assume* that the lower polarization of the $S(1)$ line is *not* caused by dilution and that most of the line photons come from shocked material within and near the outflow, these photons must be polarized by grains either within the flow itself and/or in the intervening molecular cloud between the flow and the observer. Therefore, for the polarization of the BN object to be much higher requires the object to be *behind* the flow and in addition there must be a comparable optical depth of aligned grains between BN and the far side of the outflow region, as between us and the outflow cavity.

Observations of CO absorption bands in IRc2 and BN by Geballe & Wade (1985) lend support to this suggestion that BN lies behind the molecular flow. Measuring CO absorption against bright continuum sources allows them to examine an outflow along the entire line-of-sight to the continuum source. The profiles for BN and IRc2 are very similar, suggesting the principal absorption components seen in BN's spectrum are also present in IRc2's spectrum. This includes a feature in the quiescent gas and another arising from the high-velocity flow. The conclusion is that BN cannot be in front of the molecular flow.

The high-resolution K-map of the central region clearly highlights IRc2 and BN as continuum sources. As noted earlier the overall pattern is complex and it is possible there are other contributing sources of the radiation. However, although there are several vectors which are not centro-symmetric to one of these sources, there are no other obvious centres of illumination. The breakdown in the centro-symmetric pattern around IRc9 is most likely due to dichroic absorption by aligned grains of the radiation from IRc9, identifying it as a continuum source.

Along the NW portion of the flow, the $S(1)$ polarization vector at 32W, 32N (coordinates in arcsec relative to BN – see Fig. 3) is aligned perpendicular to the flow direction, in the manner of a centro-symmetric vector around Peak 1. At (40E, 16S) the vector is rotating from being aligned with the flow to being centro-symmetric about Peak 1 indicating that scattering of radiation into the line-of-sight is important. Although the data are limited, we infer from these points that we are observing the limits to which the flow extends. Since the percentage polarization is small, the $S(1)$ flux received from these points need not all be due to scattering and could contain a component due to emission from the outflow. Nevertheless the outflow cannot extend much beyond this point, and certainly by (50W, 40N) the molecular hydrogen emission appears to be all due to reflection. At an assumed distance of 480 pc for the OMC-1 complex, this limits the projected distance of the flow from IRc2 (at 6E, 7S, Lester *et al.* 1985) to about 17 000 AU in the SE lobe (vector at 40E, 16S) and 26 000 AU in the NW lobe (vector at 32W, 32N), and almost certainly to less than 35 000 AU (vector at 50W, 40N).

One particularly interesting feature of the inner region is the apparent twist in the $S(1)$ and continuum polarization pattern in the IRc2 region, the presumed source of the outflow. Data

here are limited of course, and we therefore defer a detailed discussion until after we have obtained further observations. We speculate, however, that a twist in the polarization pattern may be due to the presence of a disc around IRc2. There may be a similarity here with twists observed in the polarization pattern around optical bipolar nebulae such as LkHα233 (Aspin *et al.* 1985) and R&T CrA (Ward-Thompson *et al.* 1985). In those objects high-resolution CCD data at optical wavelengths show the centro-symmetric patterns due to reflection of radiation from a central source except close into the source itself There the polarization is non-zero with a twisted band of aligned polarization vectors running across the source. Models to explain this involve dust discs around the source and alignment of grains by wound-up magnetic fields in the discs, with the polarization vectors lying along the disc. In the case of Orion the twist appears to be along the outflow axis and not along the disc. There are interesting consequences for the geometry of the cloud collapse and the collimation of the outflow if this effect is verified.

Even without invoking the specific effects of a disc around IRc2 there is evidence for a twisting in grain alignment in OMC-1. Lonsdale *et al.* (1980) measured large circular polarizations (1.6 and 0.7 per cent) in both BN and IRS2 (near IRc9), respectively. They argue that the high ellipticities are consistent with a model in which the polarization is produced by a medium of aligned grains with a twist in the alignment. It is possible that the twisting arises due to the presence of a dichroic sheet local to each continuum source and rotated with respect to the sheet extending throughout the outflow region.

In addition, although polarization by dichroic absorption would appear to dominate the central region, some scattering into the line-of-sight of both line and continuum radiation will occur. Since both the continuum radiation (from BN and IRc2) and the line radiation are aligned along the outflow axis, any scattered radiation, observed along this axis, will have a polarization vector orthogonal to that produced by dichroic absorption. This will lead to both a depolarization and a rotation in the position angle of polarization.

4.2 OUTER REGION

Away from the sources of continuum and line radiation the polarization appears to be dominated by scattering from dust grains. This is evidenced by the large levels of polarization and the geometry of the vector pattern.

That scattering is efficient at 3.8 μm (WDC) is a strong indication that dust grains are large in OMC-1. Rouan & Leger (1984) have proposed a model in which one of the continuum clumps, IRc4, is illuminated by IRc2. The model is self-consistent in terms of grain properties, geometry and radiative transfer. Accounting for both the fluxes and the high levels of polarization observed at K (this paper) and L' (WDC) constrains the average grain size to a range of 0.4–0.6 μm. This result is also consistent with estimates based on the ratio of extinction at 9.7 and 3.8 μm in front of IRc2. Such an estimate for grain size is considerably larger than the mean value of 0.08 μm found by Mathis (1979) for the interstellar medium.

Jets of molecular hydrogen emission have been reported by Taylor *et al.* (1984) extending from the region of IRc9 NW along the outflow axis. This is in the region of centro-symmetric reflected $S(1)$ emission. Our data would suggest that all the emission observed in this region is merely reflected from Peak 1 and not due to intrinsic emission in the line-of-sight unless it occurs on a size-scale much smaller than our beam size.

The actual extent of the dichroic sheet beyond the outflow region is not easy to determine since its effect will only be obvious when a self-luminous source is viewed directly through it, e.g. around BN, IRc2 and IRc9. The centro-symmetric pattern in the outer region cannot be due to the scattering of radiation off aligned elongated grains; this would lead to polarization vectors aligned with the long axes of the grains (Martin 1978). The scattering in the outer regions must

therefore take place off spherical particles or randomly aligned elongated particles. However, this does not necessarily limit the extent of the dichroic sheet if it lies between us and the scattering grains. Passage of the scattered radiation through aligned grains will produce distortions in the polarization, including the conversion of linear to circular polarization, but the distortions will be small provided that the polarization produced by dichroic absorption is small compared to that produced by the initial scattering. This is the likely situation in the outer regions where the scattering angles are large.

Interestingly, we are unable to determine from our data whether the continuum radiation we observe between BN and Peak 1 of the $S(1)$ emission is scattered from aligned elongated grains. Since the long axes of the grains responsible for the polarization of the line radiation are orthogonal to the outflow direction, the polarization vectors produced by scattering of continuum photons from either BN or IRc2 will be in the same direction as for spherical or non-aligned elongated grains. We do know, however, that the polarization produced by scattering of the continuum photons in this region must dominate over the polarization that might be produced by the subsequent passage of the scattered radiation through aligned grains.

4.3 COMPARISON WITH 10 μM POLARIZATION

Aitken *et al.* (1985) measured the 8–13 μm polarization of BN, IRc2, 3, 4 and a point 17 arcsec SE of BN. With the exception of IRc2 all the sources show evidence of dichroic absorption by aligned grains and have position angles directed towards the region of IRc2. The authors explained the lack of polarization for IRc2 and the orientation of the E-vectors by grain alignment produced by transfer to the grains of the intrinsic angular momentum of photons coming from IRc2. Grains will line up with their angular-momentum vectors pointed away from IRc2. The minimum energy configuration the grain can then adopt is with its principal axis of inertia along the angular momentum direction. Hence the grains will be aligned with their long axes perpendicular to the radius vector from IRc2. This mechanism therefore leads to E-vectors directed radially toward the source of photons and also accounts for the low observed polarization of IRc2.

We need to consider two points. First, the E-vectors of the 10 μm polarization of IRc4 and $S(1)$ polarization in the same region (as viewed on the sky) are orthogonal. If both polarizations are produced by dichroic aborption then there needs to be two sets of aligned grains. (There is no evidence from the results of Aitken *et al.* that the 10 μm polarization of IRc4 is produced by emission from aligned grains.) One possibility is that the 10 μm polarization is produced in a separate dichroic slab local to IRc4.

Secondly, the low 8–13 μm polarization of IRc2, despite the strength of the silicate feature, is difficult to explain if, as we suggest, the central region is covered by a dichroic slab. Also IRc2 can be seen at L' and we note from the polarization map of WDC that it has a continuum polarization of about 10 per cent. Although low levels of polarization could arise from the cancellation of polarization produced in emission by aligned warm grains (close to IRc2) and the preferential absorption by overlying cold dust (Dennison 1977), the effect could only be observed in a single band and not when a band is spectrally resolved. This is because the level of polarization seen in emission is largely independent of optical depth but the level of polarization resulting from absorption is approximately proportional to optical depth. Since the optical depth for IRc2 is strongly wavelength dependent (because of the observed silicate feature) then cancellation of polarization is not possible at all points within the 8–13 μm band.

4.4 IMPLICATIONS FOR THE KINEMATICS OF THE OUTFLOW

One of the clearest statements that can be made on the basis of the results presented here and in WDC is that at both 2.2 and 3.8 μm scattering is highly efficient, as seen by the large levels of

Astron. Astrophys. 144, 220–226 (1985)

ASTRONOMY AND ASTROPHYSICS

CCD observations of bipolar nebulae

II. Lk Hα 233

C. Aspin[1], I. S. McLean[2], and M. J. McCaughrean[1]

[1] Department of Astronomy, University of Edinburgh, Edinburgh, Scotland
[2] Royal Observatory, Blackford Hill, Edinburgh EN9 3HJ, Scotland

Received June 27, accepted October 19, 1984

Summary. Imaging polarimetry and photometry in three colours (B, V, and R) are presented of the emission star/nebula Lick Hα 233. These new observations were made with the Royal Observatory, Edinburgh CCD Imaging Spectropolarimeter and have a pixel resolution of ~0".5. The polarization maps clearly reveal a large disk or torus (probably with aligned grains) surrounding the central star and show that the associated faint nebulosity (or lobes) is part of the system. These "bipolar" type lobes are highly polarized indicative of an optically-thin reflection nebula. Polarization mapping is shown to be a powerful tool in detecting disks and bipolar lobes. We also discuss the origin of the large visual extinction towards Lk Hα 233 and conclude that an intrinsic component of extinction must be present.

Key words: polarization – bipolar nebulae – photometry – emission stars

1. Introduction

In Paper I of this series (Aspin and McLean, 1984) we presented the results of a new study of the bipolar nebula M 2-9 using a versatile CCD camera capable of photometry and polarimetry. The combination of a polarization map and a photometric image having the same spatial resolution is a useful diagnostic tool in investigating the nature of bipolar nebulae of all types. In particular, the presence of polarization indicates not only an outer reflection nebula but it can also allow one to detect and study the central disk. Continuing with this series we present the first CCD data on the peculiar object Lk Hα 233.

At position $\alpha = 22^h32^m30^s$, $\delta = +40°20'0"$: 1950, Lk Hα 233 is the 233rd entry in the Lick Observatory Hα survey published by Herbig (1960). Associated with this emission-line object are faint nebulous clouds whose appearance on photographic plates is that of a diffuse letter X centred on a bright core. On the best plates, faint outer nebulosity is also visible. Markarian et al. (1977) included it in a catalogue of extra-galactic objects since it was diffuse and showed Hα emission. Named Markarian 914, he described it as a moderately condensed interacting galaxy having a starlike nucleus. Spectroscopic observations by Taniguchi and Tamura (1982) revealed radial velocities of $< 40\,km\,s^{-1}$ and so firmly dismissed the extragalactic origin of Lk Hα 233.

Those optical spectra, and the ones obtained by Calvet and Cohen (1978) and Cohen and Kuhi (1979) showed a monotonic increase in Hα and [S II] emission in the stellar region between 1974 and 1980. Hα emission was also found in the nebula, yet other emission lines such as [O III] 5007 were absent in contrast to the situation found in many morphologically similar objects. The strong continuum present in the nebula spectrum was assumed to be due to reflection from the central star which was previously classified as spectral type A 7 by Herbig (1960). Some 2.56 mag of visual extinction were implied by the continuum structure, a value similar to that of two other objects in the same field (Calvet and Cohen, 1978). Both Lk Hα 233 and the other field objects are apparently superimposed or as Herbig (1960) reported, embedded in a dark cloud. Both Calvet and Cohen (1978) and Taniguchi and Tamura (1982) assumed that the 2.56 mag of extinction was an effect of the interstellar medium.

The only published polarimetry of Lk Hα 233 are broad-band (B, V) single-aperture measurements of the stellar component itself reported by Vardanyan (1979). He found Lk Hα 233 to be highly polarized for a stellarlike object and presented values of $P(B) = 6.9\%$ at 146° and $P(V) = 10.8\%$ at 147°. To provide a way of distinguishing reflection from emission nebulosity and to help in probing the sources of extinction, we therefore obtained CCD observations at high *spatial* resolution in three wavelength regions to provide both imaging polarimetry and photometry of the entire nebula.

2. Instrumentation and observations

All the measurements presented here were made using the Royal Observatory, Edinburgh CCD Imaging Spectropolarimeter (ISP) system which is described in detail elsewhere (see McLean et al., 1980).

Briefly, the cooled CCD is the detector in an imaging system designed to measure both flux and polarization over a spatial extent of a few arcminutes in discrete wavelength passbands in the range 3800 Å to 1 μ. For this study, a thinned back-illuminated CCD manufactured by RCA was used which has an array of 320 × 512 pixels each 30 μ square and a readout noise of 56 electrons. Since the performance of our instrument in the imaging polarimetry mode has been demonstrated previously (see McLean et al., 1983; Aspin and McLean, 1984) no further discussion will be given here.

The CCD observations of Lk Hα 233 were made on the 8th and 10th October 1983 (UT) on the 3.8 m UK Infrared Telescope on Mauna Kea, Hawaii. Observing conditions at UKIRT were clear and stable throughout the above period with seeing-limited images ≲1".5 FWHM. Using the wide-field imaging polarimetry

Send offprint requests to: C. Aspin

mode of observation (see McLean et al., 1983) we took a sequence of four exposures per filter at different orientations of a half-wave plate with a fixed Glan prism as the polarization analyser. A focal plane aperture defined an unvignetted field of view of approximately 100" in diameter (at UKIRT) with an image scale of 0".49 per pixel. To compensate for small guiding drifts the four data images needed to calculate the Stokes parameters and produce one polarization map were registered by computer interpolation to a small fraction (~ 2%) of a pixel. Flat fields were obtained by out-of-focus observations of the inside of the diffusely illuminated telescope dome.

Polarization maps were made in three wavelength ranges (*B*, *V*, and *R*) defined by interference filters; the central wavelengths and passbands (FWHM) of our filters are *B*(4600 Å/730 Å), *V*(5400 Å/700 Å), and *R*(7600 Å/772 Å). Total integration times of 12 min per waveplate position were used for each filter, giving a total observing time of 144 min for the set of three filters. Photometric calibration standards were taken from Lundolt's catalogue of Selected Area (SA) equatorial stars (Landolt, 1973) with the *R* magnitudes reported in Kunkel and Rydgren (1979) and Moffett and Barnes (1979).

Preliminary data reduction was performed at the telescope to assess the quality of the measurements but final image processing was carried out at Edinburgh using software developed by CA for the STARLINK VAX 11/780 computer network.

3. Results

Figure 1 contains four *V* filter images of LkHα 233 and its remarkable nebula at increasing contrast levels (increasing *A* to *D*). This representation is very useful in emphasizing the progression of brightness levels in the object. Note how weak the diffuse nebula is compared with the central source and how one-sided the nebula appears. The sharply curved cut-off on the left is just the edge of the focal plane diaphragm. Figure 2 shows another four images also in the *V* filter this time containing the surface brightness (*S*), the percentage polarization (*P*), the polarized flux (i.e. the product *PS*) and the polarization error image (*E*). In the latter image each point gives the polarization error expected due to photon counting statistics at that point in the real image. The corresponding polarization vector map, in which the polarization is represented by line segments, is presented in Fig. 3 overlaid on a *V* filter contour map of detected signal. This vector map is plotted at only one third of the full resolution of our CCD camera. The vector separations are approximately 1".5 with each polarization value being calculated from one 0".49 CCD pixel. A polarization vector map at the full resolution (bins spaced by 0".49) for the central region is given in Fig. 4.

On close study of Figs. 1 and 2 and Figs. 3 and 4 several interesting features can be seen. Over most of the nebula the polarization is very high (30% to 60% in all three passbands).

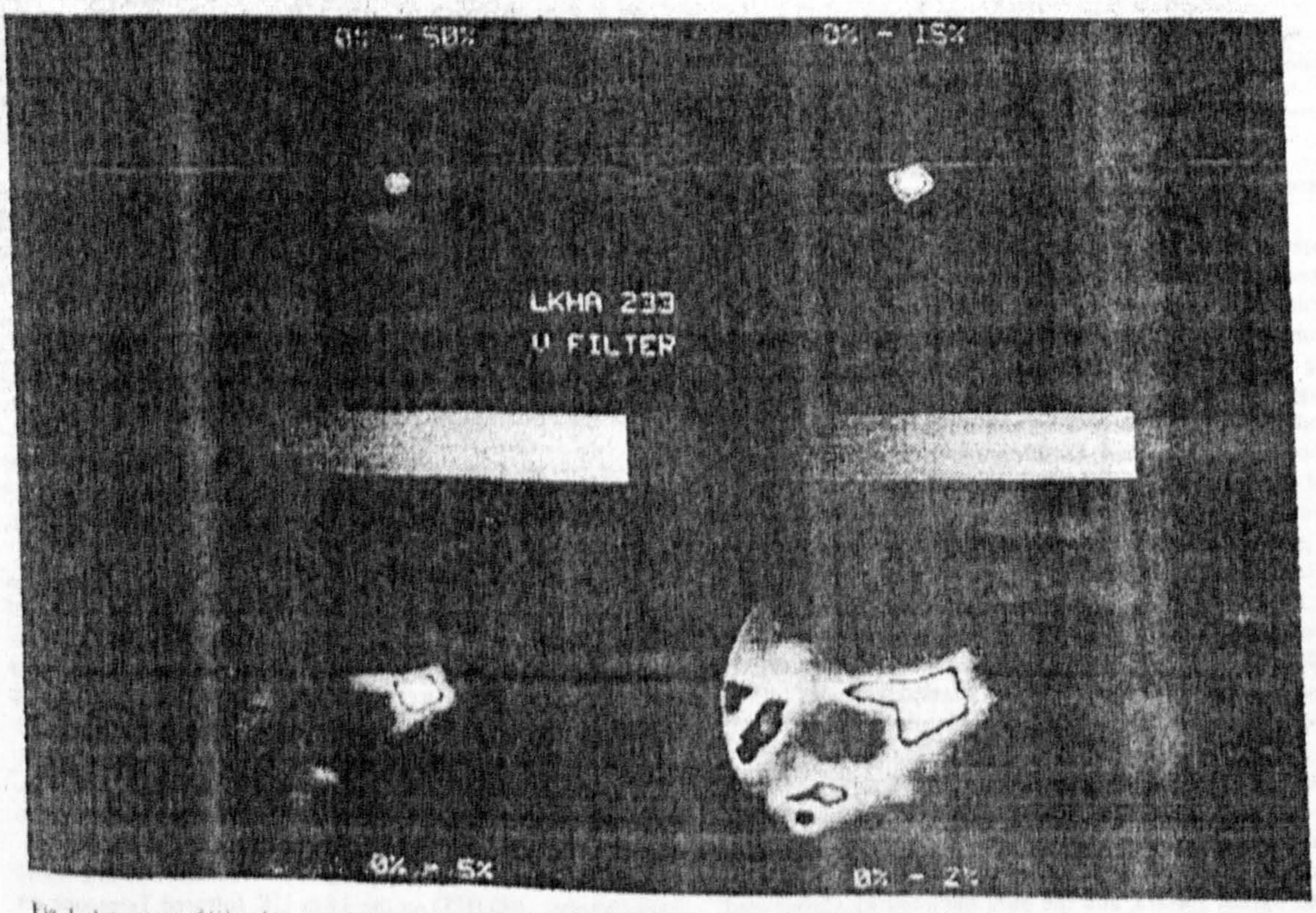


Fig. 1. A montage of *V* band images of LkHα 233 at four different contrast levels to illustrate the progression of surface brightness in the object. Image 1A (top left) shows the stellar regions and the local nebulosity. Image 1B (top right) indicates the brighter fingers of nebulosity close to the star (and illustrates the X shaped appearance seen on previous photographs). Image 1C (bottom left) shows these fingers extending into the diffuse nebula. Image 1D (bottom right) contains the whole object detected in our CCD images. The lower brightness threshold level is at 22.3 mag per square arcsecond. The nebulosity to the extreme West has a surface brightness of 21.7 mag per square arcsec in this *V* filter

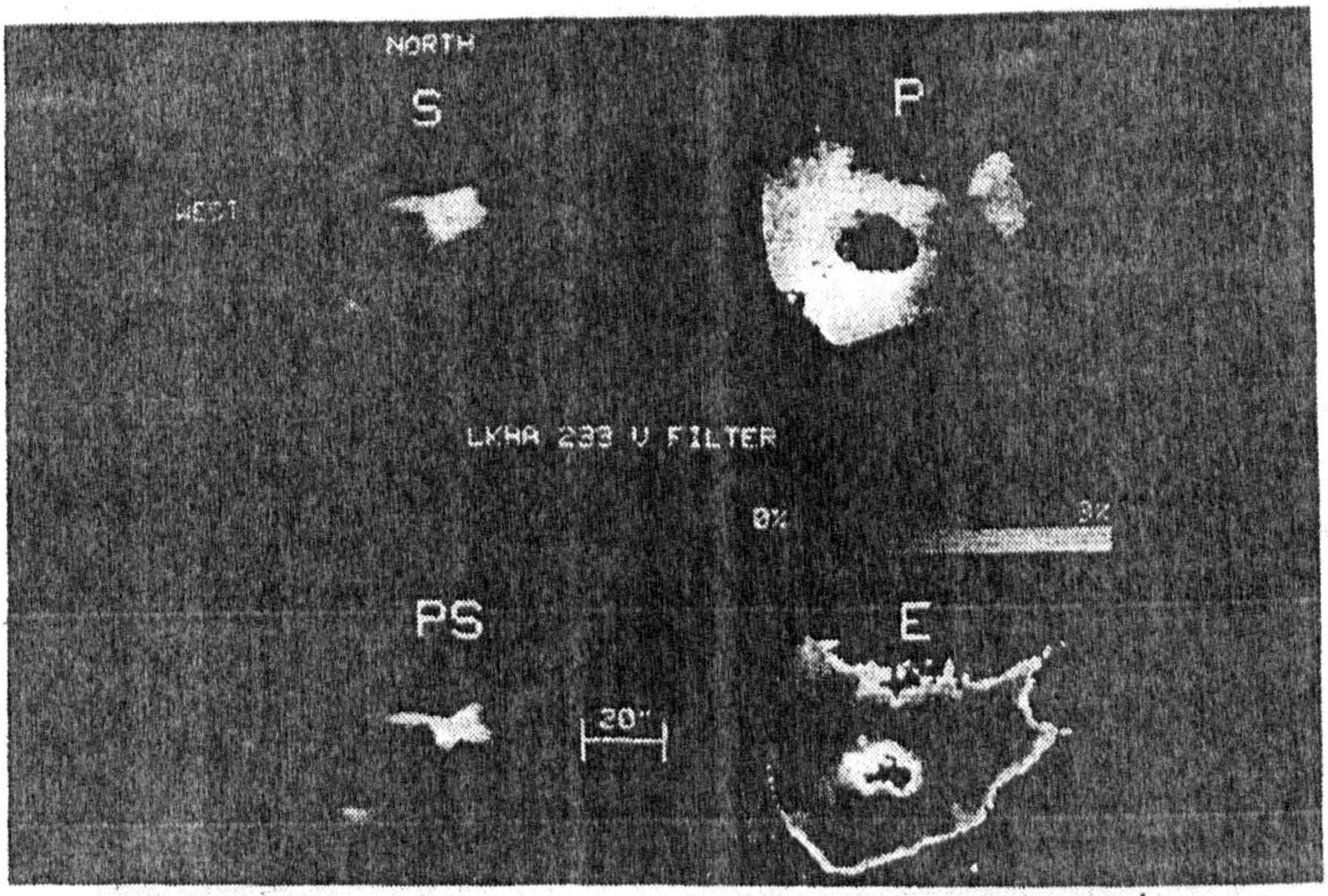


Fig. 2. This shows the surface brightness map (*S*) together with the distribution of percentage linear polarization (*P*), the polarized flux (*PS*) and the polarization error image (*E*) over the nebula. The grey scale above the *E* image gives the relationship between the percentage error in *P* and the associated grey level. Note that the central disk in the system is prominently visible in the *P* but not in the *S* image

However, a region of *low* polarization occurs to the NW and SE of the central intensity maximum (see the *P* image in Fig. 2, and the vector pattern in Fig. 4) and is seen to "fan out" from this position with percentage polarizations predominantly in the range 5% to 10%. Close to the central object lie four fingers of bright nebulosity (forming an *X*-shape). These peculiar extensions are highly polarized (30% to 35% in *V*) and become even more apparent in polarized flux (*PS* in Fig. 2). It is also noticeable that they do not appear to converge to a single point, but rather where they disappear they are separated by some 4" along the NW/SE line; this is particularly noticeable in the dark central parts of the *P* image on Fig. 2. The NW finger is significantly brighter than the others extending further from the centre before fading to the level of the diffuse nebula.

The vector map of Fig. 3 is extremely informative and shows the characteristic centrosymmetric (CS) polarization pattern of a simple reflection nebula over much of the object. This pattern confirms conclusively the speculation that LkHa 233 is the illuminating star of the associated nebulosity. However, the vector pattern deviates strongly from its centro-symmetric form in the regions of low polarization just NW and SE of the central star – the dark lane in the *P* image. In these fan-shaped regions the vectors are more nearly parallel to an axis passing through the centre of the two fans and through the central star. Also, the integrated polarization of the central source is non-zero. This indicates a real physical difference between the fan-shaped regions and the general, highly polarized nebulosity and could be geometrical in nature and/or involve a change in the particle polarizing properties. Further discussion of this remarkable effect is deferred to Sect. 4.

As a whole, the object comprises:

a) A bright central stellar core [$m(V) = 14.2 \pm 0.1$, $B-V = 0.61$, and $V-R = 1.1$ in a 5" aperture] embedded in bright nebulosity.

b) Somewhat fainter extensions/clouds to the West [$m(V) = 21.7 \pm 0.1$, $B-V = 0.45$, and $V-R = 1.05$ all per square arcs] including the regions referred to by Tanaguchi and Tamura (1982).

c) Fainter nebulous extensions to the East and North-East which rapidly decrease in brightness away from the stellar core.

To the west we have detected nebulosity out to the edge of our circular entrance aperture (i.e. to approximately 80" from the star) while to the east the nebula is not detectable further than 20" from the star.

Only the *V*-filter polarization maps are presented in full since the form of the *B* and *R* filter data is essentially identical. Note that over most of the nebulosity the polarization error due to photon statistics is $< \pm 2\%$ whereas in the brighter central regions a value of $\pm 0.5\%$ is achieved; see the *E* image in Fig. 2.

4. Discussion

On the basis of its morphology Calvet and Cohen (1978) suggested that LkHa 233 be classified as a "bipolar" type nebula. This classification is consistent with several characteristics of this object which have emerged from our CCD results. In particular, the

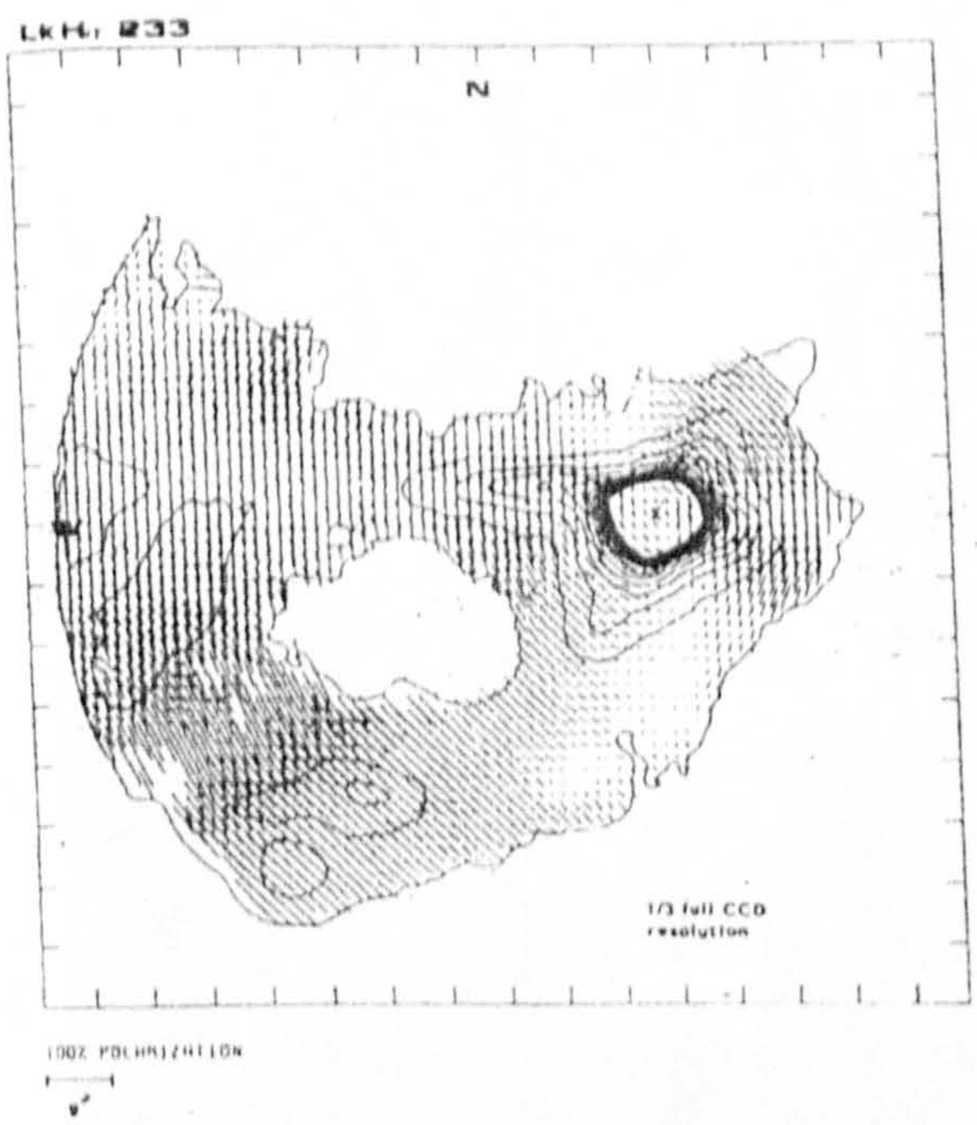


Fig. 3. A linear polarization vector map of Lk Hα 233 in the *V* filter passband. In this map the length of the vectors represents the amount of polarization with the orientation of the vector being parallel to the local *E* vector. The vectors are plotted at only *one third* (i.e. every 1".5) of the full CCD resolution. An intensity contour map of the object is overlaid with the lowest level corresponding to 2 mag fainter than the night sky. North is up and East is to the right

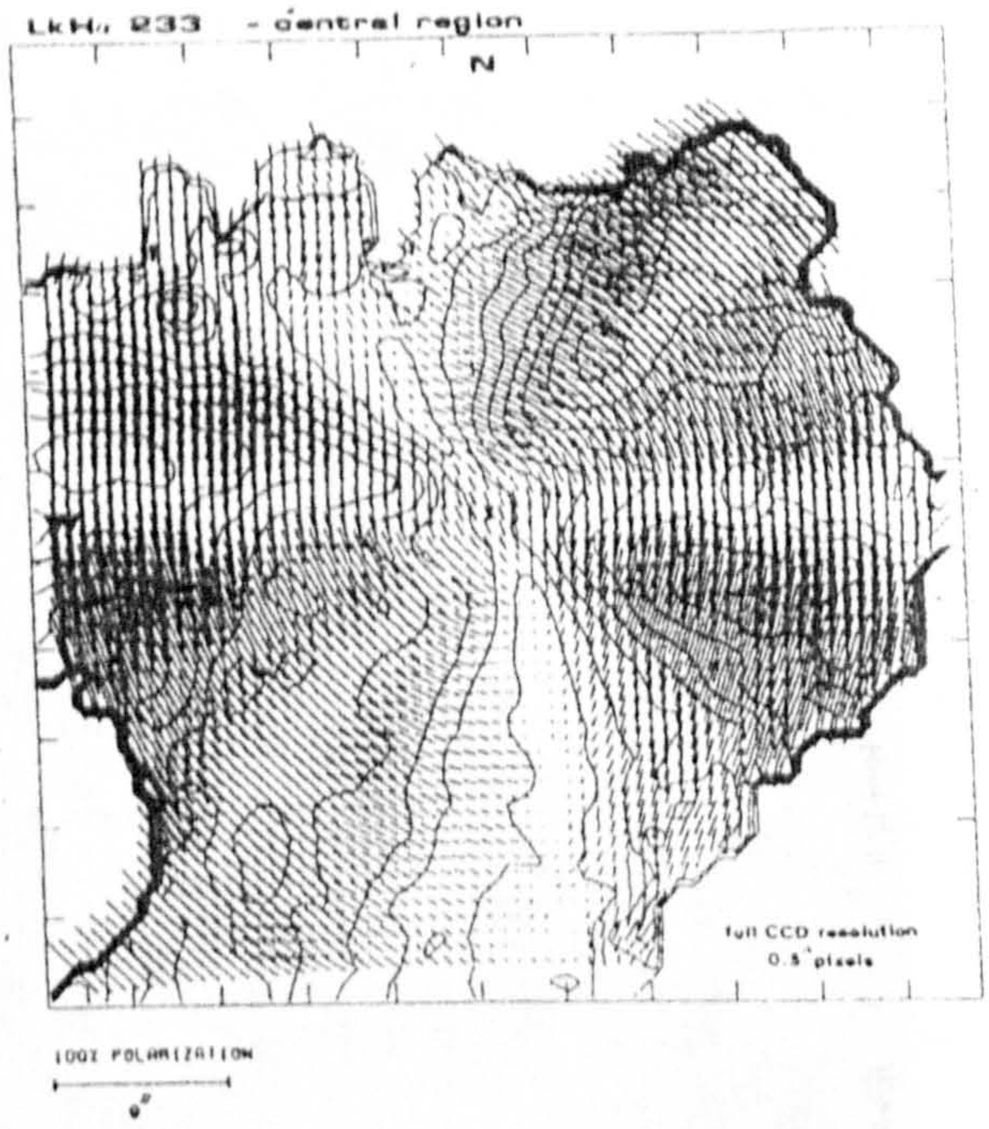


Fig. 4. A polarization vector map at the full resolution of the CCD (i.e. ~0".5 per pixel) of the central regions of Lk Hα 233 in the *V* filter. This time the contour overlay shows the percentage linear polarization *P*. The "disk" referred to in the text is clearly visible in the contours and by the disruption of the centro-symmetric pattern of vectors

polarization maps and photometric data give direct evidence for the existence of a disk of material in the Lk Hα 233 system. Actually, inspection of the intensity images alone (Fig. 1) does not really reveal a disk-like feature. In percentage polarization however, (*P* of Fig. 2) the dark fan-shaped regions of low polarization to the NW and SE of the central star are manifest. The simplest interpretation is that these regions are the cross-section of an extensive, dusty disk or torus which surrounds the central star, and thus the nebulous regions to the East and West are part of the bipolar lobes of the system.

To obtain polarization position angles orientated along the plane of the disk (i.e. perpendicular to the direction expected from the centro-symmetric pattern) implies large scale alignment of non-spherical grains within the disk itself or some very peculiar illumination geometry. If a magnetic field aligns the particles in the disk the radiation transmitted through the disk would become polarized with an orientation parallel to the field. Thus a magnetic field lying in the plane of the disk would produce polarization vectors in the sense which we observe. We have also discovered a similar effect in NGC 2261 (R Mon) which we discuss in Paper III of this series. If our interpretation is correct then it implies that in a collapsing protostar environment (with a circumstellar disk around the star) the creation of such a toroidal magnetic field may be quite common. Other grain alignment mechanisms which can be considered are those associated with mass outflow or radiation driven stellar winds. In any case our new results indicate that more detailed modelling of the physical environment around such stars is needed.

If we look at the percentage polarization *P* and the colours $(B-V)$ and $(V-R)$ along the disk (see Fig. 5) we see that the spatial behaviour is complex. At the centre of the disk, where the stellar component lies, *P* is large in all three passbands (*B*, *V*, *R*) with values of 9.6%, 9.9%, and 12.2% at position angles 145°, 147°, and 147° respectively. Our *B* value is different from the Vardanyan (1979) result of 6.9% although our position angle is in good agreement. The disk colours i.e. $(B-V)$ and $(V-R)$ are both strongly red. At increasing distances from the star along the disk *P* decreases to about 3% whereas $(B-V)$ at first becomes bluer and then becomes strongly red. This flip in the colour gradient takes place at a distance of about 10" on both sides of the central star and may be caused by nebular emission-line contamination. The variations in the $(V-R)$ colour curve are much smaller in amplitude than in $(B-V)$ but are still highly reddened values. These colour variations must be associated with the structure of the disk (i.e. its density and composition) and the central star. The blueing near the centre can be caused by the scattered blue continuum of the central A7 star. As we look further out into the disk however, the colours are dominated by the successively increasing reddening of the light before and after scattering.

The trend in the wavelength dependence of *P* in the disk together with that of the star and the diffuse nebula is shown in Fig. 6. Whereas the polarization in the star and the two regions of diffuse nebulosity rise monotonically into the red, the values in the disk (to the NW and SE) at first decrease (from *B* to *V*) and then rise to the red. From the dependence of polarization on wavelength toward the star and in the diffuse nebula we can deduce that

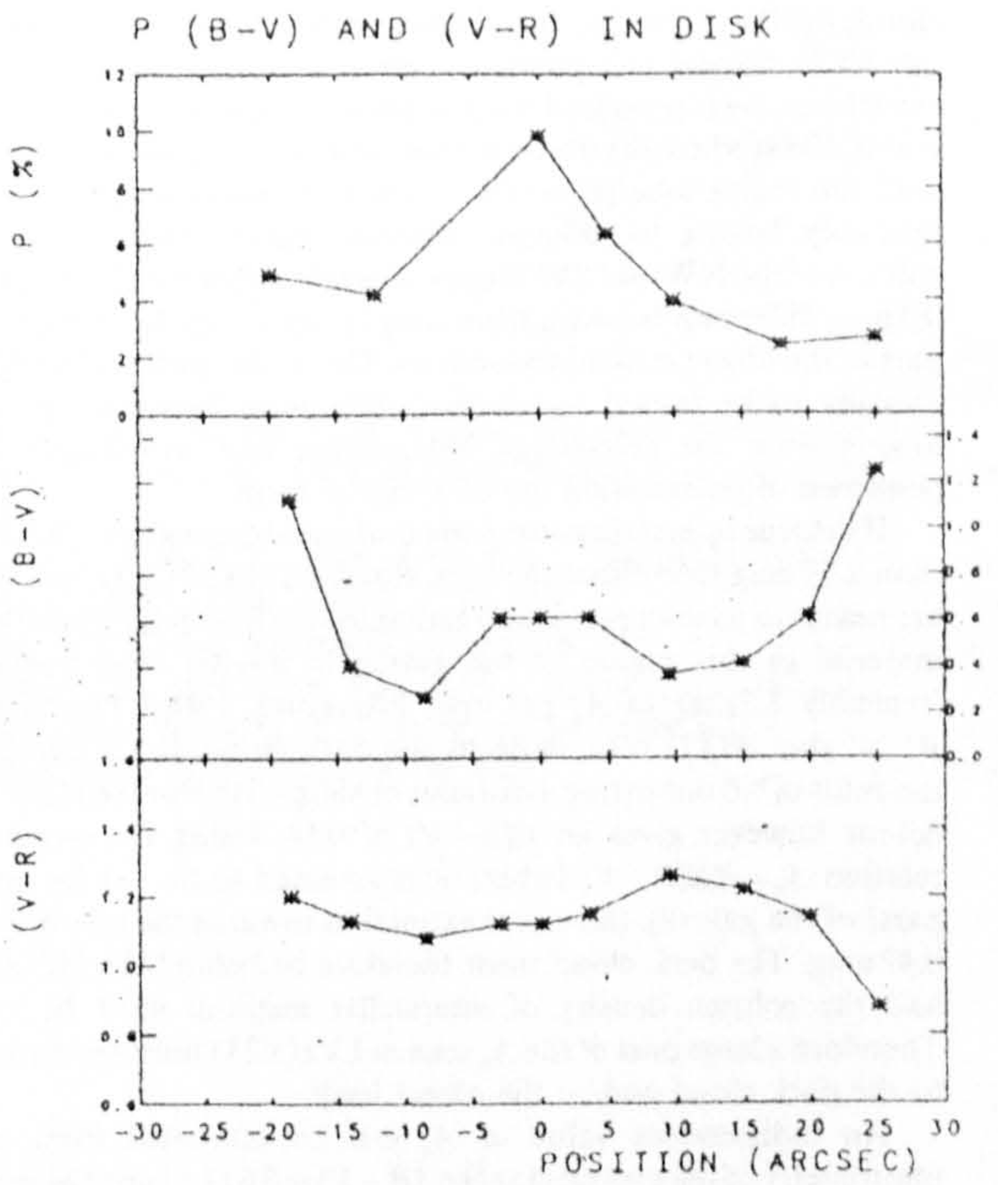


Fig. 5. The variation of percentage polarization and the $(B-V)$ and $(V-R)$ colours with position along the disk. The stellar component is situated at position 0". Each plotted value is calculated from a 5" × 5" integration bin to reduce the error. Values are given out to 20" on each side of the central star. Photon statistical errors in P are less than $\pm 0.2\%$ at all points. $(B-V)$ and $(V-R)$ colours have errors approximately ± 0.15 mag

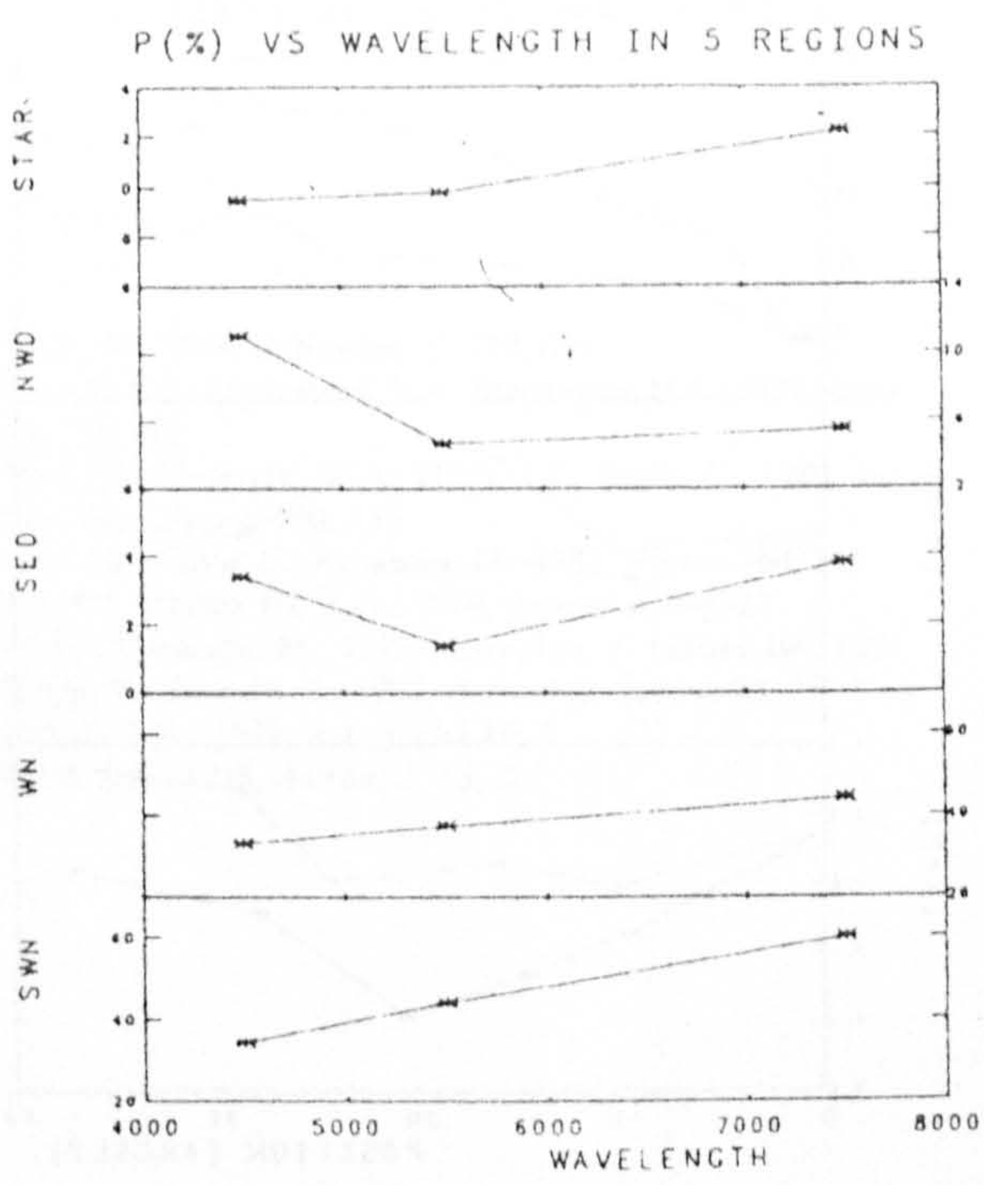


Fig. 6. The wavelength dependence of polarization, from B (4600 Å) to R (7600 Å), of the central star (STAR), the disk (NW and SE sections NWD, SED) and the diffuse nebula (W and SW regions WN, SWN). The rising polarization values into the red are indicative of scattering by large particles (radii $\gtrsim 0.07\,\mu$). Photon statistical polarization errors are of the order of $\pm 0.4\%$ or better

scattering by relatively large particles (Mie scattering) is occurring and that some foreground nebulosity is present near the star. A similar wavelength dependence was found for CRL 2688 (The Egg nebula) by Shawl and Tarenghi (1976). They calculated the expected wavelength dependence of polarization for a bipolar type model and fitted this to the CRL 2688 data. The model involved both graphite and silicate particles of sizes from 0.05 to 0.1 μ. If we compare their models with the results shown in Fig. 6 we find that only graphite particles of radius ~0.07 μ fit the polarizations in the diffuse regions and in the central star with any success. Although a monotonic increase in P with λ is seen in the above model the slope of this increase is too shallow to match our results. In the disk itself, polarization from pure graphite grains can model the increase from V to R but it fails completely to match the observed increase from V to B. We suggest a two component particle size distribution with perhaps smaller (more Rayleigh-like) particles creating the increase to the blue.

The extremely large polarizations (~60%) in the outer nebula suggest that it is 1) optically thin, 2) contains a very limited range of scattering angles and 3) is not tilted out of the plane of the sky by more than 30°. The uniformity of polarization and surface brightness in the diffuse regions (in all three passbands) implies a simple geometry and a high degree of uniformity in particle densities.

Close to the disk and central star are the four fingers of bright nebulosity referred to earlier. These are highly polarized (i.e. significantly higher than the disk and slightly higher than the part of the nebula they are projected against) and show a wavelength dependence similar to the diffuse nebula. Figure 7 shows the $(B-V)$, $(V-R)$ colours and the polarization along the bright NW finger and the fainter one to the SE. It is clear from the diagram that the behaviour of the two colours and the polarization is the same along both, although it seems that the SE finger is redder by up to 0.3 mag in both $(B-V)$ and $(V-R)$ than the other. The percentage polarization is considerably smaller in the SW finger but still has the same spatial dependency. Such differences could arise if scattering occurs in the forward direction in the NW finger and in the backward direction in the SW finger. The observed colour and polarization of the two fingers may then be attributed to an orientation effect with respect to the plane of the sky.

Another unusual feature of Lk Hα 233 is the fact that no extensive nebulosity is visible to the East beyond ~20". On inspection of earlier photographs (especially Markarian et al., 1977) faint nebulosity is seen to the East but there is still considerably less than to the West. It is true to say therefore that there must be a significant brightness difference between the two lobes. We can place an uppwer limit of 22.5 mag/square arcsecond in V for the surface brightness of the East lobe of the nebula. This means that at a distance of ~40" from the star the eastern lobe is fainter than the West lobe by at least 1 mag in V. Also, there appears to be a lack of nebulosity in the centre of the West lobe i.e. roughly along the axis of symmetry of the object. This could be caused by material being ejected only from the stellar regions along the bright fingers. Alternatively, the absence of material in the centre of the lobes could be a projection effect such that we only see the bright rim of conical lobes due to the larger column density

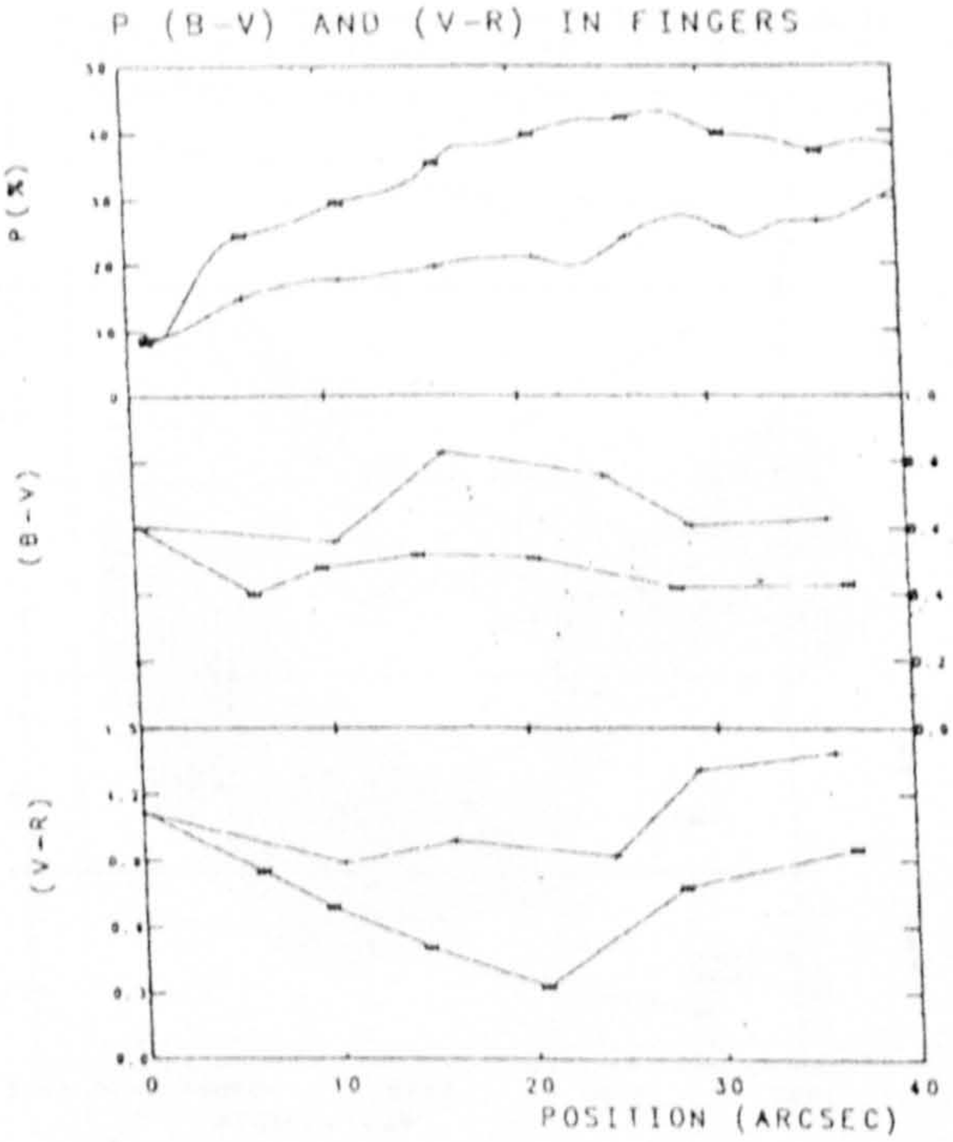


Fig. 7. The variation of P, $(B-V)$, and $(V-R)$ along both the NW (•) and SW (+) fingers of bright nebulosity in Lk Hα 233. It can be seen that the SW finger is redder than the NW finger by at least 0.3 mag in both $(B-V)$ and $(V-R)$ along its whole length. This finger is also of lower polarization than its NW companion even though it has the same spatial structure. Values of $(B-V)$ and $(V-R)$ are again calculated from 5" integration bins. Errors in P, $(B-V)$, and $(V-R)$ are similar to those quoted for Fig. 5

of scatterers along the edges of the nebula; in other words, a shape consistent with biconical outflow. Since the polarizations seen are extremely high in the diffuse regions it seems reasonable to assume that the latter explanation is correct.

From spectroscopic observations Calvet and Cohen (1978) calculate that 2.56 mag of visual extinction, A_v, is present towards Lk Hα 233. This was assumed to be of interstellar origin. Supporting this assumption is the fact that two other Hα emission-line objects in the same field (Lk Hα 231 and 232) have similar values of A_v (2.35 and 2.44 mag respectively). If this conclusion is correct, then it is unusual that there is no intrinsic component of A_v in Lk Hα 233, especially in the light of our discovery of a substantial amount of dust material enveloping the central star. Examining the position of Lk Hα 233 in relation to Lk Hα 231, 232 and the dark cloud in which all the objects are believed to be embedded, it appears that in projection Lk Hα 231, 232, and 233 appear at similar positions with respect to the cloud i.e. near the centre of a dark band which forms part of the larger cloud complex. A significant amount of extinction could be created within the dark cloud itself. In fact, the apparent agreement between the A_v's of all three objects could be fortuitous in the sense that Lk Hα 231 and 232 may be more deeply embedded in the dark cloud and the intrinsic component of extinction in Lk Hα 233 compensates for its "surface" location. Such a configuration could also account for the apparent absence (or severe dimming) of the East lobe in 233 if the system is tilted so that the East lobe is embedded deeper in the dark cloud; the East lobe may then be below the detection threshold of our CCD images and previous optical photography. A similar model has been proposed for the peculiar nebula PV Cep (Levreault, 1984) where the object is near the edge of a molecular cloud with the visible lobe protruding from it. These objects however, probably belong to different physical classes. Also, from the colours of the NW and SW fingers of bright nebulosity (Fig. 7) the 0.3 mag difference between them may be accounted for if the NW part of the object is tilted towards us. This additional reddening is unlikely to be caused by physical differences between the two fingers since the percentage polarization and wavelength dependence of polarization are of a similar form.

If the true interstellar component of visual extinction is smaller than 2.56 mag then either the dark cloud, Lk Hα 231, 232, and 233 are nearer to us than previously estimated or the column density of material in this region of the galaxy is smaller than average (typically 2.5 mag of A_v per kpc; Fitzgerald, 1968). The bright $B1.5V$ star HD 213976, close to the dark cloud, has a distance modulus of 9.6 indicating a distance of 880 pc. Its observed $(B-V)$ colour however gives an $E(B-V)$ of 0.14. Using the standard relation $A_v = RE(B-V)$ (where R is assumed to be ~3 for most parts of the galaxy), the visual extinction towards the star is only 0.42 mag. The dark cloud must therefore be behind the HD star and the column density of interstellar medium must be low. Therefore a large part of the A_v seen in Lk Hα 233 must be *intrinsic* to the dark cloud and/or the object itself.

An independent value of A_v can be estimated from our photometry of the central star [i.e. $(B-V)=0.61$]. Using the usual value of $R=3$ and taking the intrinsic $(B-V)$ value of an A7 star (0.25) an A_v of only ~1 mag is obtained. Taking a value of R implied by the photometric properties of HD 213976 would make this smaller still. Since our B filter is centred on a longer wavelength than the standard B filter it may be more prone to contamination by emission-line flux, thus affecting the $(B-V)$ colour and hence the calculated extinction value. If we use $(V-R)$ colour instead and, from the work of Johnson (1966), take an intrinsic $(V-R)$ for an A7 star as 0.15 we find an $E(V-R) \sim 0.95$. Using the extinction law of Whitford (1958) i.e. $A_v = 2.5E(V-R)$ we find an $A_v \sim 2.33$ mag in better agreement with the results of Calvet and Cohen (1978).

5. Conclusions

We have presented the first CCD polarization maps and photometry of the emission-star nebula Lk Hα 233 in three passbands (B, V, and R). These new data establish the existence of a large circumstellar disk or torus of aligned grains and confirm that the bipolar lobes are seen in reflection. It is the polarization observations that yield the cleaest evidence for the central disk of material. Relatively large dust grains of size $\lesssim 0.1\,\mu$ are responsible for most of the polarization seen in the lobes. It is postulated that the orientation/position of the object with respect to its associated dark cloud may be responsible for the apparent morphology of the system and that Lk Hα 233 may be a young stellar object.

Acknowledgements. It is a pleasure to acknowledge the support of UKIRT Staff. We wish also to thank the referee for useful comments on parts of this paper. We gratefully acknowledge the SERC for research fellowship/studentship support of CA and MJM respectively, and thank PATT for the allocation of telescope time.

References

Aspin, C., McLean, I.S.: 1984, *Astron. Astrophys.* **134**, 333
Calvet, N., Cohen, M.: 1978, *Monthly Notices Roy. Astron. Soc.* **182**, 687
Cohen, M., Kuhi, L.V.: 1979, *Astrophys. J. Suppl.* **41**, 743
Fitzgerald, M.P.: 1968, *Astron. J.* **73**, 983
Herbig, G.H.: 1960, *Astrophys. J. Suppl.* **4**, 337
Johnson, H.L.: 1966, *Ann. Rev. Astron. Astrophys.* **4**, 193
Kunkel, W.E., Rydgren, A.E.: 1979, *Astron. J.* **84**, 633
Landolt, A.U.: 1973, *Astron. J.* **78**, 959
Levreault, M.: 1984, *Astrophys. J.* **277**, 634
Markaryan, B.E., Lipovetskii, V.A., Stepanyan, D.A.: 1977, *Astrophys.* **13**, 215
McLean, I.S., Cormack, W.A., Herd, J.T., Aspin, C.: 1980, *Soc. Phot. Inst. Engng.* **290**, 155
McLean, I.S., Aspin, C., Reitsema, H.: 1983, *Nature* **304**, 243
Moffett, T.J., Barnes III, T.G.: 1979, *Astron. J.* **84**, 627
Shawl, S.J., Tarenghi, M.: 1976, *Astrophys. J. Letters* **204**, L25
Taniguchi, Y., Tamura, S.: 1982, *Astrophys. Letters* **23**, 25
Vardanyan, L.A.: 1979, *Astrofizika* **16**, 3
Whitford, A.E.: 1958, *Astron. J.* **63**, 201

Is the polarization of NGC1068 evidence for a non-thermal source?

I. S. McLean*, C. Aspin†, S. R. Heathcote† & M. J. McCaughrean†

* Royal Observatory, and † Department of Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK

NGC1068 is one of the brightest galaxies included by Seyfert[1] in his list of extragalactic objects having compact, luminous nuclei within which broad, high-excitation emission lines occur. It has been the subject of intensive studies at UV[2,3], optical[3-6], IR[7-12] and radio wavelengths[13-15]. Unresolved questions concern the nature and relationship of the sources of the excess flux seen in the UV and IR, their connection with the collimated jets apparent in high-resolution radio maps and their association with the extended region responsible for the broad emission lines. A further question is the location of any dust and its role in modifying the optical and UV spectrum. We report here observations with two high-resolution optical spectropolarimeters which throw new light on these questions. From detailed structure found in the linear polarization spectrum of the nucleus we conclude that dilution by starlight modifies the polarization to an extent not previously appreciated. In fact, the polarization of the non-stellar flux in the optical and near IR is approximately independent of wavelength (as expected for synchrotron emission or electron scattering) with a direction orthogonal to that of the radio jets; such an arrangement is reminiscent of certain quasars and radio galaxies[16,17].

Optical polarization in NGC1068 was first reported by Walker[18] who found that the observed polarization rose strongly into the blue. Subsequently, Visvanathan and Oke[19] attempted to account for the observed flux and polarization as a superposition of an unpolarized stellar component and a non-stellar component with polarization independent of wavelength. However, it was not until the work of Angel *et al.*[20] that spectropolarimetry with enough resolution to separate the effects of lines and continuum was obtained. Those measurements had a variable resolution (20Å at 3,300Å; 160Å at 6,500Å) and were sufficient only to demonstrate general trends. Accurate broad-band polarimetry[20] revealed a weak (5°) rotation of position angle (θ) with wavelength (λ) across the optical region. In addition, a small component of circular polarization was detected (~0.2%). It was concluded[20,21] that these results favoured a dust-scattering polarization mechanism.

Recently, we have observed NGC1068 using two new multichannel spectropolarimeters. One of these is an imaging spectropolarimeter (ISP) with a charge coupled device (CCD), and the other is a Pockels cell polarimeter used in conjunction with the spectrograph and Image Photon Counting System (IPCS) at the Anglo-Australian Telescope. Both of these polarimeters (designed and built at the Royal Observatory Edinburgh) are described elsewhere[22,23].

The observations of NGC1068 reported here were made on the 3.9-m Anglo-Australian Telescope. Those with the CCD system were obtained on 29 August and measurements with the Pockels cell/IPCS were taken on 2 September 1982. The entrance apertures used were 1.7×2.2 arc s for the CCD and 2.0×2.5 arc s for the IPCS. Instrumental calibrations have been applied using standard techniques[22,23]; the IPCS data are photon noise-limited whereas the CCD data contain some extra noise due to CCD defects and charge shifting losses. Figures 1 and 2 display the ISP and Pockels cell data, respectively.

One feature of the new data is that the $p(\lambda)$ curve is not at all smooth but exhibits a complex structure of bumps and dips which are not related to the emission lines which dominate the spectrum. Of particular interest is the region from λ4,500 to λ5,300 Å because this interval is covered by both instruments. By comparing Figs 1 and 2 it is quite evident that the broad area (or notch) of low polarization around 4,750 Å has been detected by both polarimeters. On re-examining the earlier low-resolution observations of Angel *et al.*[20] we noticed that the '4,750 Å notch' and several other ripples were indeed present but were apparently mistaken for noise.

There is also structure associated with the emission lines themselves. In our data, the [O III] lines at 4,959 and 5,007 Å are resolved. Across both lines there is a strong depolarization and a very large (~45°) rotation of θ. In the Pockels cell data (see Fig. 2) there is perhaps evidence for substructure in p and θ within the 5,007-Å line. There may also be structure in p associated with the [S II] lines and the [N II]/Hα blend (Fig. 1). Because we observe the flux and polarization we can derive the polarization of the emission lines by themselves by interpolating for the continuum polarization in the line. Using the Pockels cell data for the strong 5,007 line the average intrinsic polarization of the emission is $p = (1.3 \pm 0.2)\%$ at $\theta = 151 \pm 4°$. This result agrees with the values obtained by Angel *et al.* for aperture sizes of 2 and 3 arc s which are $(0.6 \pm 0.2)\%$ at $143 \pm 8°$ and $(1.0 \pm 0.1)\%$ at $149 \pm 4°$.

For the permitted lines the situation is ambiguous because of the complexity of the continuum p (λ) curve. The Hα line and [N II] lines are blended, Hβ and He II (4,686) lie in the 4,750 Å notch and Hγ is weak. Nevertheless, it seems likely that part of the decrease in p within the 4,750-Å notch may be due to unpolarized He II and Hβ emission.

In addition to the continuum and line structure already mentioned, our data reveal another new feature, namely relatively narrow enhancements in p just redward of the Balmer emission lines. As the Hβ feature is recorded by both polarimeters we are confident that it is not an instrumental effect. The feature is broadened by binning in the Pockels cell data and by the instrumental profile in the CCD data. These enhancements in p do not correlate with any obvious intensity structure in the red wings of the Balmer lines but resemble displaced 'mirror images' of the emission lines themselves. The redshift of these

0028-0836/83/330609—03$01.00

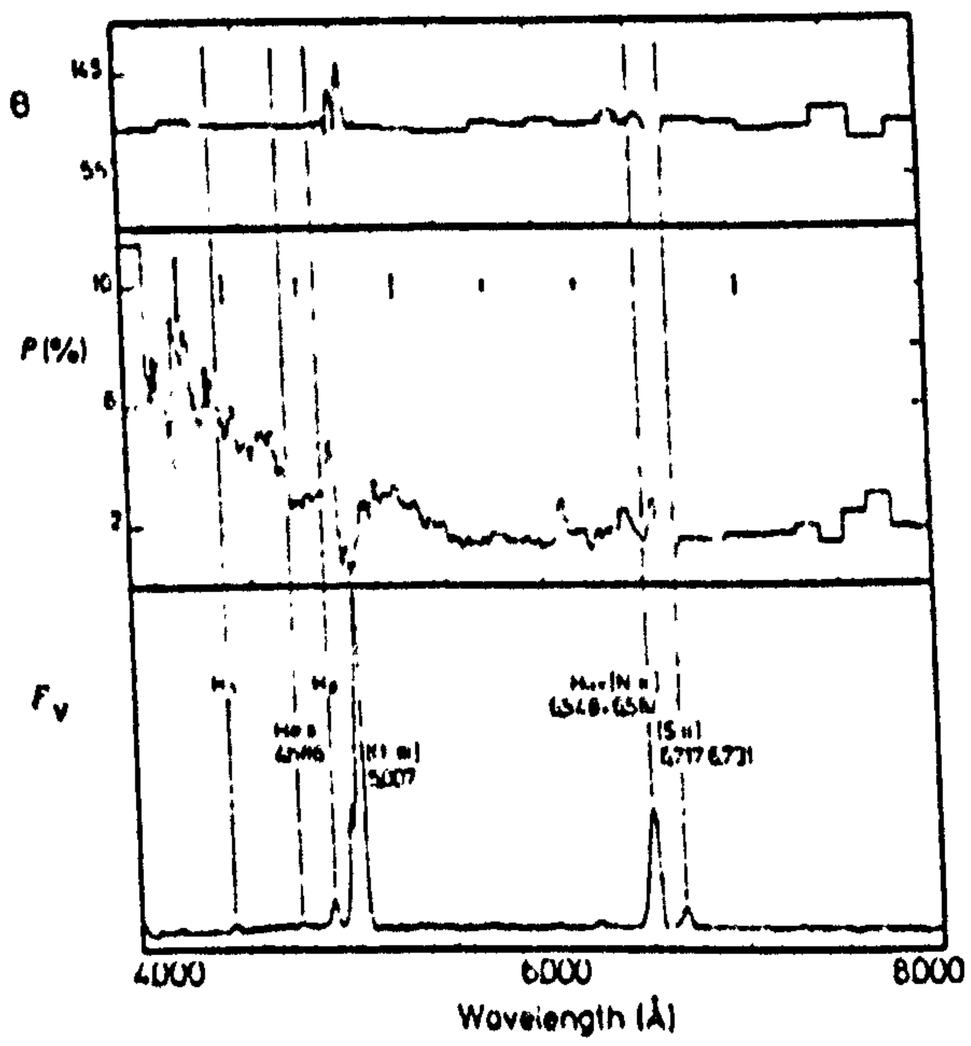


Fig. 1 CCD imaging spectropolarimeter data showing the position angle (θ), the percentage linear polarization (p) and flux (F_ν) as a function of wavelength. Typical error bars per bin in p are shown; the error in θ is 28.65° σ_p/P. The bin width is 15 Å over most of the p data; the resolution is ~30 Å. A small segment around λ 6,700 Å has been deleted due to a group of bad pixels.

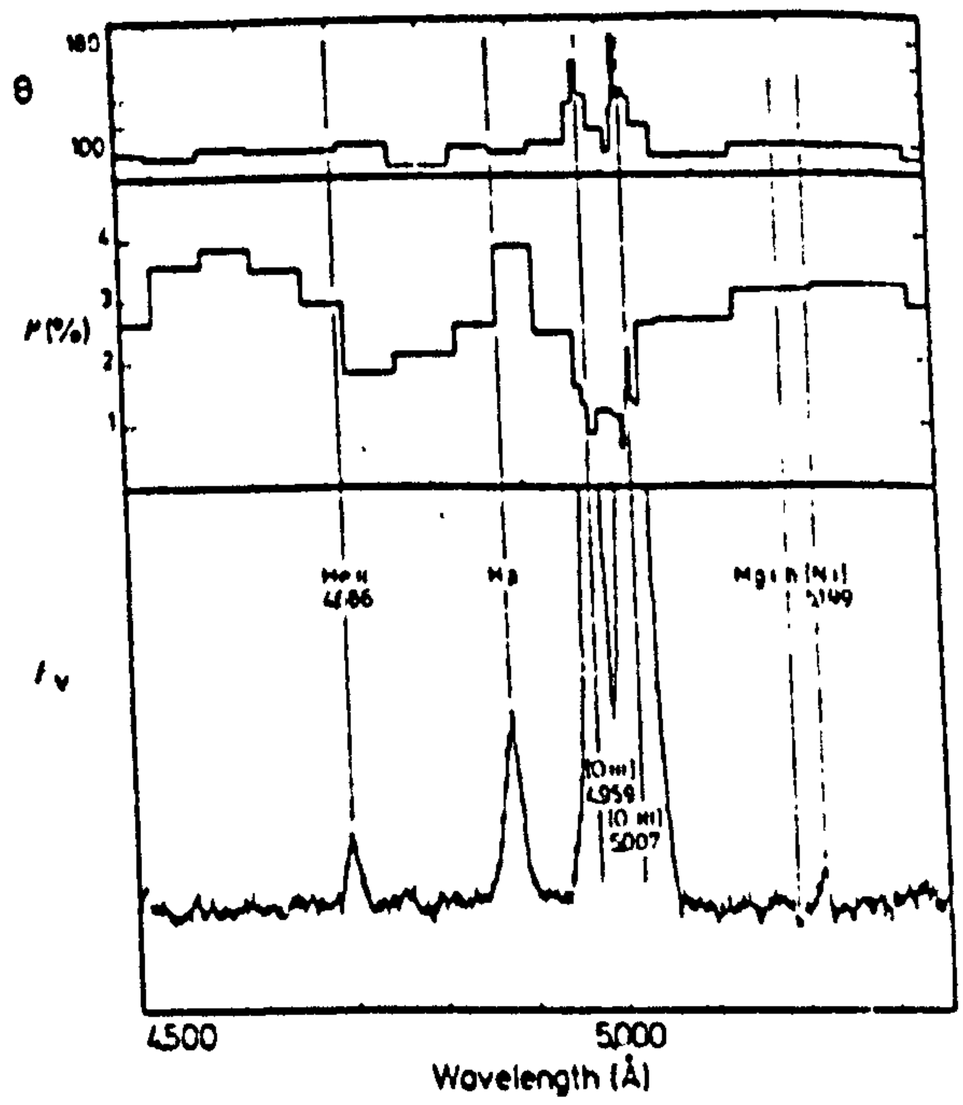


Fig. 2 Pockels cell/IPCS data. Position angle (θ), percentage linear polarization (p) and flux (F_ν) are shown as a function of wavelength. The data have been binned such that the expected photon noise error in p is 0.5% outside the [O III] lines and 0.4% within the lines. At the full resolution a single bin corresponds to 0.5 Å.

features relative to the Balmer line cores is ~50 Å (that is ~3,000 km s^{-1} at 1Hβ).

To interpret these observations we first need to consider qualitatively the continuum polarization. Any underlying polarized continuum arising in the nucleus of NGC1068 must inevitably be diluted by the integrated light of stars in that part of the galaxy in the line of sight. Therefore, the observed polarization should show largest values (least dilution) at wavelengths corresponding to stellar absorption lines. For a normal spiral galaxy the spectrum is dominated by solar and late-type stars, and there are absorptions due to *H* and *K* bands of Ca (the 4,000 Å break), the *G* band of CH (4,300 Å), the Mgb (5,174 Å) feature, the Balmer lines of hydrogen and molecular bands of several species of molecules. In addition, the decline of the stellar spectra towards short wavelengths will result in less dilution and therefore more polarization in the blue.

From observations of the galaxy itself[16] the position angle of the rotation axis is 145°. This is close to the direction of the polarization of the [O III] lines. Polarization due to transmission through grains aligned by a galactic magnetic field is expected to produce polarization vectors parallel to the galactic plane (that is, $\theta \sim 55°$ not 145°). On the other hand, polarization of the [O III] lines could be due to scattering into our line-of-sight by grains on the outskirts (farthest from the nucleus) of the emitting regions. On average, these emitting regions seem to be aligned with the projection of the major axis of the galaxy. Because dilution by galactic light will affect the emission lines less than the neighbouring continuum the permitted lines must in fact be substantially less polarized than the continuum, contrary to the findings of Angel *et al.* The absence of any obvious change in θ does, however, indicate that they do not share the polarization of the forbidden lines.

Since dilution leaves the plane of polarization invariant it cannot account for the weak wavelength dependence of θ revealed by broadband observations unless the galaxy light is itself slightly polarized at a different angle from the underlying source. Such polarization can readily occur by passage through the interstellar medium of NGC1068. Because the galactic flux is weakest in the far blue, the observed position angle of 102° should correspond most nearly to that of the non-stellar component. A direction of 102° is almost orthogonal to the projected direction of the radio jet, once allowance is made for its curvature[14,15]. We estimate the jet to lie in position angle ~11° in the optical nucleus whereas the elongated central radio component itself would give ~20°; this orthogonal geometry also pertains further out in the jet as demonstrated by 15-GHz polarization data[14]. If the polarized emission is interpreted as being due to the synchrotron process then the magnetic field is aligned along the jet.

The enhancements in p associated with the Balmer lines are puzzling. One explanation may be that they correspond to strong redshifted absorption lines. At the wavelengths of these absorption lines dilution would be considerably reduced and so the polarization would tend to rise. As far as we know, no set of redshifted absorption lines has been detected in the spectrum. Alternatively, they may be due to cloud–cloud scattering, perhaps in high-velocity jet material. For example, a photon emitted from a cloud moving in one direction could be scattered by a cloud moving in the opposite direction and thus suffer a shift in wavelength. The dispersion of velocities observed in NGC1068 is almost high enough to be consistent with such a model (~2,000 km s^{-1}).

To quantify the idea that dilution has a crucial role we have constructed a simple, semi-empirical model for the continuum. The underlying non-stellar (nuclear) component was taken to be a power law $F_\nu \propto \nu^{-\alpha}$ with an intrinsic polarization p_∞ independent of λ (as is appropriate for optically-thin synchrotron emission from a homogeneous source or for electron scattering) at a position angle perpendicular to the radio jet so that $100° < \theta_\infty < 120°$. The spectrum of the galactic component was taken to be the same as that of the nucleus of M31[24] (which, like NGC1068, is of type Sb). The interstellar polarization of the galactic component was assumed to follow Serkowski's empirical relationship[25], which is characterized by two parameters λ_{max} and p_{max} (the wavelength of maximum polarization and its value). We adopt a position angle for the polarization

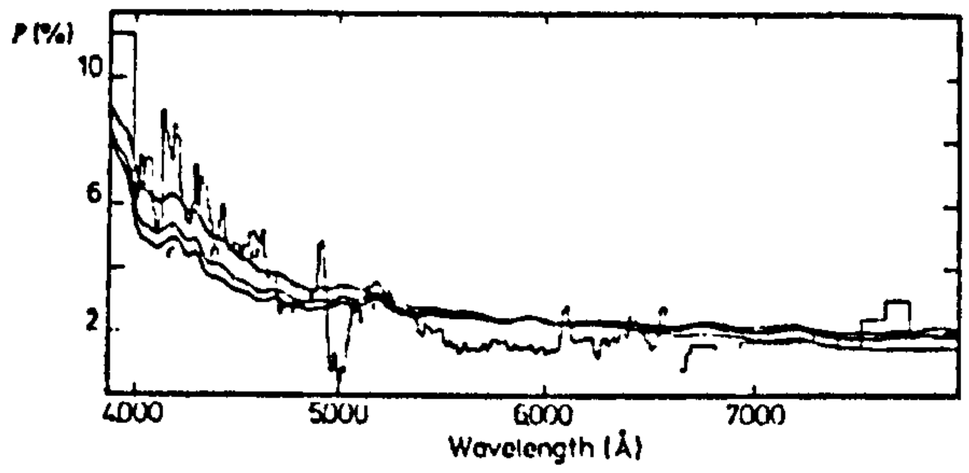


Fig. 3 Representative results from the continuum polarization model described in the text. For all three $p(\lambda)$ curves the interstellar contribution in NGC1068 is given by $p_{max} = 0.6\%$, $\lambda_{max} = 5{,}000$ Å, $\theta_{gal} = 55°$; the position angle of the non-stellar component is $\theta_{ns} = 102°$; the template galaxy is M31 and both the non-stellar and the galactic components suffer an extinction $A_v = 0.4$ mag. The remaining parameters for each curve are as follows. Model *a*: $\alpha = 0.0$, $F_{ns} = 0.23$, $P_{ns} = 12\%$; Model *b*: $\alpha = 1.0$, $F_{ns} = 0.20$, $P_{ns} = 13\%$; Model *c*: $\alpha = 1.5$, $F_{ns} = 0.166$, $P_{ns} = 15\%$. Model *a* gives the best fit to our optical data while model *c* reproduces the IR results.

of the galactic component $\theta_{gal} \sim 55°$ as expected for extinction by grains aligned by a magnetic field lying parallel to the galactic plane of NGC1068. To specify the relative contributions of the two components we define F_{ns} as the fraction of total light contributed by the non-stellar component at 5,500 Å within a circular aperture 2 arc s in diameter. Either or both components may be reddened; the wavelength dependence of the extinction (A_v) is assumed to follow the average relationship for our galaxy[26]. Figure 3 presents the results of some representative models superimposed on the ISP data.

Values of λ_{max} and p_{max} are constrained by the weak $\theta(\lambda)$ dependence in the continuum and F_{ns} is (for a given α) constrained by the need to match the observed spectrum. The principal free parameters are therefore α and p_{ns}. Acceptable fits to our data have $0 < \alpha < 1.5$; larger values produce too small a variation in p between 4,000 and 5,000 Å while smaller values cause too rapid a decline into the red. To constrain the model parameters further it is desirable to use IR observations. It seems[12,27] that the trends seen in the optical polarization reverse beyond 1 μm so that p increases to a broad maximum (~4.2% in a 4 arc s aperture) near 2.2 μm, while θ rotates back to larger angles. This is also a dilution effect as normal galaxies peak around 1.6 μm. To produce such an upturn at all, however, requires values of α in the upper part of the range given above. In fact, the model having $\alpha = 1.5$ yields $p = 1.6\%$, 2.4% and 4.4% at 1.25, 1.65 and 2.2 μm respectively, in reasonable accord with the results of Lebofsky *et al.*[12] but is not so satisfactory in the optical. A value of $\alpha \sim 1.0$ is a plausible compromise. The departure from the model in the 5,200–6,500Å range may indicate that the synchrotron source is actually inhomogeneous. In common with other studies[2–4], the values of A_v in our model are very much smaller than those found for the emission line regions themselves[3,12,28,29]. This would seem to imply that much of the dust lies within the emission-line clouds.

The values of α and F_{ns} found here are in good agreement with those obtained by Koski[4] from modelling of the intensity spectrum alone. M. Ward (personal communication) has estimated $F_{ns} \sim 30\%$ from published aperture photometry. Recent estimates based on a comparison of the strengths of stellar absorption lines in the blue with those in normal galaxies[30,31] imply $F_{ns} \sim 70\%$. With such a large value of F_{ns} neither a simple synchrotron model nor a pure scattering model is tenable but an inhomogeneous synchrotron source cannot be ruled out.

A model invoking pure dilution of synchrotron emission must account for the observation of a small but significant component of circular polarization. Normal synchrotron emission could only account for this circular polarization if the magnetic field were very strong, $\sim 10^3$ G. Two alternative explanations can be advanced, namely, multiple scattering in dust clouds (which would also give a large linear polarization due to scattering) or single scattering of a highly linearly polarized beam through a medium of aligned grains[32,33]. The latter mechanism may be efficient in NGC1068 because the difference of position angles between synchrotron polarization and grain alignment is near to the optimum (that is, 45°).

We thank the staff of the AAO for support, and our colleagues Malcolm Smith, Adrian Webster and Ray Wolstencroft, Martin Ward and David Axon for their comments. This work was supported by the SERC.

Received 26 April; accepted 16 June 1983.

1. Seyfert, C. K. *Astrophys. J.* **97**, 28–40 (1943)
2. Boksenberg, A. *et al. Nature* **275**, 404–414 (1978)
3. Neugebauer, G. *et al. Astrophys. J.* **238**, 502–509 (1980)
4. Koski, A. T. *Astrophys. J.* **223**, 56–73 (1978).
5. Shields, G. A. & Oke, J. B. *Astrophys. J.* **197**, 5–16 (1975).
6. Penston, M. V., Penston, M. J., Selmes, R. A., Becklin, E. E. & Neugebauer, G. *Mon. Not. R. astr. Soc.* **169**, 357–393 (1974).
7. Becklin, E. E., Mathews, K., Neugebauer, G. & Wynn-Williams, C. G. *Astrophys. J. Lett.* **186**, L69–L72 (1973)
8. Telesco, C. M., Becklin, E. E. & Wynn-Williams, C. G. *Astrophys. J. Lett.* **241**, 169–172 (1980)
9. Hildebrand, R. H. *et al. Astrophys. J.* **216**, 698–705 (1977).
10. Cutri, R. M. *et al. Astrophys. J.* **245**, 818–828 (1981)
11. McCarthy, D. W., Low, F. J., Kleinmann, S. G. & Gillett, F. C. *Astrophys. J. Lett.* **257**, L7–L11 (1982).
12. Lebofsky, M. J., Rieke, G. H. & Kemp, J. C. *Astrophys. J.* **222**, 95–99 (1978)
13. de Bruyn, A. G. & Willis, A. G. *Astr. Astrophys.* **33**, 351–356 (1974)
14. Wilson, A. S. & Ulvestad, J. S. *Astrophys. J.* **263**, 576–594 (1982).
15. Pedlar, A., Booler, R. V., Spencer, R. E. & Stewart, O. J. *Mon. Not. R. astr. Soc* **202**, 647–656 (1983).
16. Stockman, H. S. & Angel, J. R. P. *Astrophys. J. Lett.* **220**, L67–L71 (1978)
17. Antonucci, R. R. J. *Nature* **299**, 605–606 (1983).
18. Walker, M. F. *Astrophys. J.* **151**, 71–97 (1968)
19. Visvanathan, N. & Oke, J. B. *Astrophys. J. Lett.* **152**, L165–L168 (1978)
20. Angel, J. R. P., Stockman, H. S., Woolf, N. J., Beaver, E. A. & Martin, P. G. *Astrophys. J. Lett.* **204**, L5–L9 (1976)
21. Nandy, K. & Wolstencroft, R. D. *Nature* **225**, 621 (1970).
22. McLean, I. S., Cormack, W. A., Herd, J. T. & Aspin, C. *Proc. Soc. Photo-opt. Instrum. Engng* **290**, 155–200 (1981).
23. McLean, I. S., Heathcote, S. R., Peterson, M. J., Fordham, J. & Shortridge, K. *Mon. Not. R. astr. Soc.* (submitted).
24. Oke, J. B. & Schwarzschild, M. *Astrophys. J.* **198**, 63–70 (1975)
25. Serkowski, K. in *Interstellar Dust and Related Topics* (eds van de Hulst, H. C. & Greenberg, J. M.) 145–159 (Reidel, Dordrecht, 1973).
26. Savage, B. D. & Mathis, J. S. *Ann. Rev. Astr. Astrophys.* **17**, 73–111 (1979).
27. Knacke, R. F. & Capps, R. W. *Astrophys. J. Lett.* **192**, L19–L22 (1974).
28. Malkan, M. A. *Astrophys. J. Lett.* **264**, L1–L6 (1983).
29. Gaskell, C. M. *Publ. astr. Soc. Pacif.* **94**, 891–893 (1982)
30. Malkan, M. A. & Oke, J. B. *Astrophys. J.* **265**, 92–106 (1983)
31. Malkan, M. A. & Filippenko, A. V. *Astrophys. J.* (submitted)
32. Marun, P. G., Illing, R. & Angel, J. R. P. *Mon. Not. R. astr. Soc.* **159**, 191–201 (1972)
33. Lonsdale, C. J., Dyck, H. M., Capps, R. W. & Wolstencroft, R. D. *Astrophys. J. Lett.* **238**, L31–L34 (1980)

0028–0836/83/330611—04$01.00

polarization in the centro-symmetric parts of the continuum and line maps. This result has important consequences for the interpretation of the kinematic structure of the high-velocity flow.

(i) Observation of line photons along a given line-of-sight does not necessarily give a true measure of the column density or velocity dispersion of the gas along that line-of-sight.

(ii) Whenever a differential velocity field exists between the reflecting grains and the gas (with stationary grains and moving gas, or moving grains and stationary gas) then a Doppler shift will be induced in the reflected light.

If such a phenomenon is occurring then the observed line profiles of molecular hydrogen will contain a significant Doppler-shifted component due to scattering and the profile cannot be used as a direct measure of the true wind velocity. Doppler shifts up to twice the differential velocity of the gas and grains can be produced in this way depending on the scattering angle.

Geballe *et al.* (1986) have recently argued against the importance of scattering as a mechanism for producing the wings on the molecular hydrogen on the basis of differences between the profile of the $S(1)$ line 2.12 μm and the $O(7)$ line at 3.81 μm. However, they incorrectly assume that there is a marked difference between the scattering efficiencies at the two wavelengths. Results presented here and in WDC argue strongly against such an assumption.

Establishing if Doppler scattering is significant for the molecular hydrogen line is of major importance if we are to determine the kinematics of the high-velocity flow. The obvious initial line of attack would be to measure the polarization as a function of velocity across the line profile of the molecular hydrogen and other lines (e.g. Brγ). If a significant increase in polarization occurs in the wings of the lines then this provides unambiguous evidence that they are mainly produced by reflected light. However, the converse conclusion is unfortunately not implied by the opposite result. Indeed, even the discovery of zero polarization in the wing of the lines cannot rule out that they are Doppler shifted by scattering. To understand this statement it is necessary to realize the distinction between the degree of polarization and the scattered flux. The former depends critically on both the level of diluting radiation and the geometry of the illuminating source. In the case of the OMC-1 continuum, for example, the illumination is highly anisotropic, being very much like a classical reflection nebula. For the molecular hydrogen the situation is likely to be far less extreme, within the worst case of cylindrical symmetry, and thus the polarization levels will be low. However, the observed Doppler shifts will be identical in both cases for illumination at a given angle θ and its complement $180-\theta$, relative to the velocity of the gas. In order to produce the net blueshift required of the scattered line photons, one simply requires that the gas is approaching the scatterers, that is the scattering grains are foreground of the intrinsic molecular hydrogen emission region.

4.5 ALIGNMENT MECHANISM

The simplest interpretation of the uniform alignment of the $S(1)$ polarization vectors parallel to the outflow axis is that of dichroic absorption of radiation in a medium of aligned, elongated dust grains. Two popular mechanisms for grain alignment are the Davis–Greenstein mechanism (Davis & Greenstein 1951) in which elongated spinning paramagnetic dust grains are aligned by a magnetic field, and photon alignment in which alignment of grains is due to the transfer of the intrinsic angular momentum of photons. For the latter mechanism this leads to the angular-momentum vectors of the grains being along the radial direction from the source (presumably IRc2) and hence the distribution of E-vectors is also radial, which is not consistent with the 2.122 μm $S(1)$ observations. For the Davis–Greenstein mechanism the angular-momentum vectors will line up along the magnetic field direction. The long axes of the grains will

orientate themselves perpendicular to this direction (the minimum energy configuration for rotational energy).

The polarization component of the transmitted radiation along the grain axis will be preferentially absorbed, and the polarization vectors will point along the magnetic field direction. In this scenario, therefore, the *polarization vectors in the outflow region are tracing the magnetic field direction*.

Other possible alignment mechanisms are the streaming of high-velocity molecular gas with respect to the dust grains (Gold 1952) and grain motions induced by radiation pressure of photons from IRc2. However, in both these cases, grain alignment will be with the long axis *along* the outflow axis. Hence the polarization vectors will be orthogonal to that observed.

Following Spitzer (1978) a comparison of the time-scale for the Davis–Greenstein alignment τ_{DG} with that of randomizing grain angular momenta through collisions τ_{COL}, and with the age of the flow τ_{FLOW}, suggest that

$$\tau_{DG} \sim \tau_{FLOW} \gg \tau_{COL} \qquad \text{for } B \sim 100\,\mu\text{G}.$$

From this we deduce that it is unlikely that such a magnetic field will be able to orientate ordinary paramagnetic grains within the flow and that grain alignment by paramagnetic relaxation is not viable if the fields are so small. Indeed as pointed out by previous authors (Johnson *et al.* 1981; Aitken *et al.* 1985) the closeness of the gas and dust temperatures within Orion makes it difficult for the mechanism to work since it operates on the difference between grain and gas temperatures.

However, paramagnetic relaxation may still be possible for the following reasons. First, the grains may be superparamagnetic, with the iron in the grains being clumped in groups of 100 atoms; this is not unreasonable for a region of recent star formation. Secondly, much larger magnetic fields (few mG) are suggested both by shock models and from measurements of Zeeman splitting in OH masers. Lastly, Johnson *et al.* (1981) have proposed that the Purcell (1975, 1979) spin-up mechanism producing suprathermal grain rotations operates in Orion by the preferential recombination and desorption of hydrogen at active sites on the grain surfaces. A requirement for the mechanism to operate is that a significant fraction of molecular hydrogen is dissociated and this could occur by the propagation of shocks in the inner region of Orion–KL. A more detailed discussion of these mechanisms is postponed until a later paper.

5 Conclusions

The polarization of the broad-band continuum at K has been mapped in the core of the Orion Molecular Cloud One with the highest spatial resolution so far, and with lower resolution in a more extensive surrounding area. In addition, the first detailed polarization maps in the $v=1-0$ $S(1)$ transition from shocked molecular hydrogen have been obtained.

The main results of this work are:

(i) Outside the outflow region of the cloud the polarization is due to reflection of photons by dust grains. Polarization values are large, exceeding 20 per cent in places. For continuum radiation the source of photons is the infrared source IRc2, and to a lesser extent BN. For the $S(1)$ photons the source is the region of peak molecular hydrogen emission ('Peak One'), displaced to the NW of the infrared sources. This observation amounts to the discovery of a molecular hydrogen reflection nebula.

(ii) In the molecular outflow region, the continuum polarization pattern is complex. It has the centro-symmetric pattern due to reflection modified close to the continuum sources by the overlying dichroic absorption. The $S(1)$ line polarization vectors are aligned with the outflow

axis. Polarization values are between 5 and 10 per cent with position angle about 120°. This can result from passage of the line radiation through a dichroic slab of material between us and the outflow. In this slab, elongated grains are orientated with their long axes perpendicular to the outflow direction. We are unable to determine the extent of the dichroic slab outside the outflow region.

(iii) There is a clear distinction along the outflow axis between the regions of aligned and reflected polarization. This probably delineates the extent of the outflow.

(iv) There is a twist in the position angle of the $S(1)$ electric vectors in the vicinity of IRc2. This may be indicative of effects from a disc around IRc2.

(v) The high efficiency of scattering at both 2.2 and 3.8 μm considerably complicates the interpretation of the kinematics of the high-velocity outflow. If there is any differential motion between grains and gas, which seems likely, then Doppler shifts will be produced in reflected light. The obvious test of this interpretation is to measure the polarization across the line profile. However, it should be clear that while the detection of significant polarization in the wings of the line will provide unambiguous evidence for the operation of this mechanism the converse result does not rule it out.

A self-consistent picture has emerged in which IRc2 is the source of a bipolar outflow and the dominant source of continuum radiation from the cloud. The BN object most likely lies behind the flow. Dust grains outside the limit of the flow, act to form a reflection nebula. Closer to the line-of-sight to IRc2 the shocked molecular hydrogen emission is extended (perhaps from the wall of a cavity) and this allows us to use it as a probe of the alignment of grains. Dichroic absorption by alignment of paramagnetic grains can explain the morphology of the $S(1)$ map, with the magnetic axis essentially parallel to the outflow axis, provided the fields are larger than in the general interstellar medium and/or that the grains are superparamagnetic.

Further extensive infrared polarization observations of this and several other objects are currently in progress. More detailed discussions and modelling will be given in a future paper.

Acknowledgments

We are grateful to the UKIRT and AAT Staff for assistance at the telescopes, particularly David Beattie for invaluable help in installing the Kyoto polarimeter at UKIRT. We acknowledge PATT for allocation of telescope time and travel funds. M. G. Burton, M. J. McCaughrean and R. P. Garden acknowledge the support of SERC through SERC Studentships. We thank the Japanese Ministry of Education for scientific support. This paper was written as part of a UK/Japanese collaboration. We also thank Ray Wolstencroft for several helpful suggestions on a draft of the paper.

References

Aitken, D. K., Bailey, J. A., Roche, P. F. & Hough, J. H., 1985. *Mon. Not. R. astr. Soc.*, 215, 815.
Allen, D., Bailey, J. A. & Hyland, A. R., 1984. *Sky Telesc.*, 67, 222.
Aspin, C., McLean, I. S. & McCaughrean, M. J., 1985. *Astr. Astrophys.*, 144, 220.
Axon, D. J. & Taylor, K., 1984. *Mon Not. R. astr. Soc.*, 207, 241.
Bailey, J. A. & Hough, J. H., 1982. *Publs astr. Soc. Pacif.*, 94, 618.
Becklin, E. E. & Neugebauer, G., 1967. *Astrophys. J.*, 147, 802.
Beckwith, S., Persson, S. E., Neugebauer, G. & Becklin, E. E., 1978. *Astrophys. J.*, 223, 464.
Capps, R. W., Dinerstein, H. L. & Werner, M. W., 1980. *Bull. Am. astr. Soc.*, 12, 865.
Capps, R. W., Gillett, F. C. & Knacke, R. F., 1978. *Astrophys. J.*, 226, 863.
Davis, L. & Greenstein, J. L., 1951. *Astrophys. J.*, 114, 206.
Dennison, B., 1977. *Astrophys. J.*, 215, 529.

Downes, D., Genzel, R., Becklin, E. E. & Wynn-Williams, C. G., 1981. *Astrophys. J.*, 244, 869.
Dyck, H. M. & Beichman, C. A., 1974. *Astrophys. J.*, 194, 57.
Dyck, H. M. & Lonsdale, C. J., 1980. In: *Infrared Astronomy, IAU Symp. No* 96, pp. 223–236. eds Wynn-Williams, C. G. & Cruikshank, D. P., Reidel, Dordrecht, Holland.
Dyson, J. E., Taylor, K., Axon, D. J. & Hughes, S., 1986. *Mon. Not. R. astr. Soc.*, 221, 155.
Erickson, N. R., Goldsmith, P. F., Snell, R. L., Berson, R. L., Huguenin, G. R., Ulich, B. L. & Lada, C. J., 1982. *Astrophys. J.*, 261, L103.
Gautier, T. N., Fink, U., Treffers, R. R. & Larson, H. P., 1976. *Astrophys. J.*, 207, L129.
Geballe, T. R. & Wade, R., 1985. *Astrophys. J.*, 291, L55.
Geballe, T. R., Persson, S. E., Simon, T., Lonsdale, C. J. & McGregor, P. J., 1986. *Astrophys. J.*, in press.
Genzel, R., Reid, M. J., Moran, J. M. & Downes, D., 1981. *Astrophys. J.*, 244, 884.
Gold, T., 1952. *Mon. Not. R. astr. Soc.*, 112, 215.
Hasegawa, T., Kaifu, N., Inatani, J., Morimoto, M., Chikada, Y., Hirakayaski, H., Iwashita, H., Morita, K., Topo, A. & Akabane, K., 1984. *Astrophys. J.*, 283, 117.
Johnson, P. E., Ricke, G. M., Lebofsky, M. J. & Kemp, J. C., 1981. *Astrophys. J.*, 245, 871.
Joyce, R. R. & Simon, T., 1982. *Mon. Not. R. astr. Soc.*, 200, 39P.
Knapp, G. R., Phillips, T. G., Huggins, P. J. & Redman, R. O., 1981. *Astrophys. J.*, 250, 175.
Lester, D. F., Becklin, E. E., Genzel, R. & Wynn-Williams, C. G., 1985. *Astr. J.*, 90, 2331.
Lonsdale, C. J., Dyck, H. M., Capps, R. W. & Wolstencroft, R. D., 1980. *Astrophys. J.*, 238, L31.
Martin, P. G., 1978. In: *Cosmic Dust*, Clarendon Press.
Mathis, J. S., 1979. *Astrophys. J.*, 232, 747.
Plambeck, R. L., Wright, M. C. H., Welch, W. J., Bieging, J. H., Baud, B., Ho, P. T. P. & Vogel, S. N., 1982. *Astrophys. J.*, 259, 617.
Purcell, E. M., 1975. In: *The Dusty Universe*, eds Cameron, G. B. & Cameron, A. G. W., Neal Watson, New York.
Purcell, E. M., 1979. *Astrophys. J.*, 231, 404.
Rouan, D. & Leger, A., 1984. *Astr. Astrophys.*, 132, L1.
Spitzer, L., 1978. *Physical Processes in the Interstellar Medium*, Wiley Interscience, New York.
Taylor, K. N. R., Storey, J. W. V., Sandell, G., Williams, P. M. & Zealey, W. J., 1984. *Nature*, 311, 236.
Ward-Thompson, D., Warren-Smith, R. F., Scarrott, S. M. & Wolstencroft, R. D., 1985. *Mon. Not. R. astr. Soc.*, 215, 537.
Werner, M. W., Dinerstein, H. L. & Capps, R. W., 1982. *Astrophys. J.*, 265, L13.
Wynn-Williams, C. G., Genzel, R., Becklin, E. E. & Downes, D., 1984. *Astrophys. J.*, 281, 172.

System design of a 1-5 μm IR camera for astronomy

I.S. McLean, T.C. Chuter, M.J. McCaughrean, J.T. Rayner

Royal Observatory, Edinburgh,
Blackford Hill, Edinburgh, EH9 3HJ, UK.

Abstract

The design of a near infrared imaging system for a large ground-based telescope is described and discussed. Solid-state hybrid detector arrays are used. The performance modelling of such devices is considered and their potential for IR astronomy is discussed.

Introduction

The Royal Observatory, Edinburgh (ROE), is presently engaged in a project to provide the 3.8 m United Kingdom Infrared Telescope (UKIRT) with the capability to obtain direct infrared images at high spatial resolution without recourse to raster-scanning of the telescope and a single element detector. In the first instance attention is being given to a 1-5μm (near infrared) imaging system called IRCAM (for Infrared Camera). At the heart of IRCAM is a solid-state 2-dimensional detector array with on-chip multiplexing enabling many thousands of detectors (pixels) to be read out very rapidly in series after the infrared image has been accumulated as photocharge on the detector. Although such systems are common in optical astronomy, useful devices are only now becoming available to the infrared community. A high resolution camera is needed to compliment low resolution maps and the IRAS all-sky survey.

Several constraints were borne in mind when specifying the system requirements and examining possible designs. Not least of these was the scientific desire to embody in a single instrument a useful range of operating modes or configurations to entice a broad spectrum of users. In addition, our aim was to provide an instrument which is relatively easy to use and consequently, much effort was devoted to a consideration of automating its operation and providing user-friendly software. Such features are especially beneficial during operation on the summit of Mauna Kea (14,000 ft).

There are, of course, constraints on the design of an infrared imager for UKIRT which are of a strictly physical nature and will obviously have an impact on the scientific uses of the camera. Undoubtedly the most important of these concerns the optical matching of the fixed pixel size of a commercially available array detector, to the intrinsic image scale of the telescope in order to achieve a certain angular resolution on the sky. For a given telescope and a given pixel size, the angular resolution on the sky (also called the pixel field-of-view) is a function only of the f-number of the final camera optic before the detector. At least initially, we intend to use the 58 x 62 InSb direct readout array, manufactured by SBRC, and which has 76 μm x 76 μm pixels. For UKIRT this implies that PFOV = 4.0/F# (arcsec).

There are many other practical constraints. For example, the total field of view and its impact on the physical size and cost of optical components, the need to cool the detector to a temperature between that of liquid helium and solid nitrogen to minimise dark current, the size and weight of the IRCAM and the resulting consequences for mechanical flexure, the data rate from the array, and so on. In the design outlined below we have tried to address these problems.

Overall system description

A simple block diagram showing the major components of the system is given in Figure 1. The array control unit, array electronics, A/D converter and data pre-processor computer are collectively called IRACS for Infrared Array Control System.

There are also several independent levels of software which are required. Our processors are programmed in DEC Micro Power Pascal and the UKIRT VAX 11/730 computer uses FORTRAN 77.

In summary the system works as follows. Infrared radiation emerging from the UKIRT Instrument Support Unit (ISU2) comes to a focus in front of the Collimator Optics Module. The diverging f/36 rays are then collimated into a parallel beam of light and redirected into the IRCAM Camera Body where the beam is refocussed by cold optics onto the detector array. Both of these units (Collimator and Camera Body) are seated on a large Instrument Mounting Platform protruding from ISU2 and braced with struts to the telescope mirror cell.

A Fabry-Perot etalon can be placed in the collimated beam in front of the IRCAM entrance window and a Polarimeter Module can be placed in the ISU2 above its dichroic beam-splitter mirror. Crucial drive electronics for the detector array together with some signal processing and digitising electronics for the output signal are located as close as possible to the IRCAM cryostat.

In a cabinet attached to the mirror cell is the Array Timing Generator/ Computer Interface and the Array Control electronics which is a unit incorporating an LSI 11-type microcomputer. The function of this micro-processor is to decode the high-level commands from the UKIRT VAX computer into procedures recognisable by the electronic drive system hardware and to synchronise the digitisation of the output signal from the array. In another cabinet is the motor control and drive unit required for the various moving parts (driven by stepper motors) such as filter wheels.

Data from the array is digitised and sent to a remote pre-processing computer (a μPDP 11/23+ also of the LSI 11 type) in which many rapidly accumulated exposures can be co-added into a single "image frame" before that frame is transferred to the UKIRT VAX 11/730 computer for final storage, analysis and display.

The astronomer interacts with the system via a terminal and a software routine which guides him through the various options (e.g. choice of filter, choice of integration time etc.). A false-colour image display system is available for visual presentation of the data frames received by the VAX computer. Once an infrared image data frame has been stored on disk by the observing program the user is free to use the Reduction and Analysis Package to manipulate and display the data in a variety of ways. Finally, the data files on disk are archived to magnetic tape in one of several standard formats.

Optical and mechanical design

In essence, the prime optical requirement is to match the f/36 image scale (1.54"/mm) of the UKIRT (effective aperture = 3.7m) to the pixels on a 2-dimensional array detector to achieve pixel fields of view on the order of 0.5 to a few arcseconds over total fields of view of 1 - 2 arcminutes. Moreover, it is essential to (i) form an image of the telescope mirror inside the cryostat to act as a cold stop; (ii) provide a parallel input beam to the cryostat to enable the camera to be used as an imaging Fabry-Perot spectrometer; (iii) minimise the number of optical surfaces because of light-loss and ghosting and (iv) accommodate a wide wavelength range.

The scheme adopted to satisfy these conditions is a variant of a classical design often employed in astronomical instruments. In this approach the f/36 beam emerging from the telescope is first collimated by an off-axis gold-coated curved mirror to produce bundles of parallel beams corresponding to different points in the image plane. The mirror is placed one focal length (F) from the telescope focal plane and has the property that it produces an image (the "entrance pupil") of the telescope entrance aperture at a distance slightly greater than F from its surface. This image forms a crucial "stop" for the system since all the collimated bundles (from any part of the field of view) pass through this image. In practice, the physical stop at this position is slightly oversized (10%) to allow for a relative shift between image and stop due to misalignment and flexure arising as the telescope moves.

To re-image the telescope focal plane a camera lens (of focal length f) is placed in the collimated beam just after the stop. Since the detector is physically small, the image in the telescope focal plane must be greatly demagnified by the camera optic. The magnification factor is just f/F. For example, for f = 61mm, F =610mm, m = 1/10; the f-number of the camera is essentially f/3.6 (or m x $F\#_{TEL}$). This design forms the basic optical configuration of IRCAM and the real components yield a PFOV = 1.2 arcseconds.

In the practical realisation of this design a second surface - that of a flat (plane) mirror has been added to "fold" the system in a compact way and minimise off-axis angles. Figure 2 shows the practical ray diagram/ physical layout; the curved and flat (gold-coated) mirrors comprise the Collimator Module, and a cold, anti-reflection-coated Zinc Selenide lens within the cryostat acts as the camera optic.

Two filter wheels each with a capacity for 10 large diameter (30mm) components are located in the collimated beam just in front of the cold stop. Standard broad band filters, a wire-grid polarizer, and narrow band (1% or 2% bandwidth) filters are included. Immediately in front of the cryostat window can be placed a Fabry-Perot etalon of the type already available at UKIRT. With the existing etalons the field of view will be limited to ~1 arcmin; larger etalons can be accommodated.

To baffle this optical system to stray radiation from warm surfaces outside the cryostat two basic steps are taken. Firstly, a cold baffle extends from the entrance window

back into the cryostat. Secondly, two reflecting annular curved mirrors are placed in the beam outside the cryostat such that they "look back" at cold surfaces.

The most significant mechanical questions to be faced are (i) how to mount the required external optics and the cryostat itself onto the telescope, (ii) how to minimise flexures, and (iii) how to provide a large enough optical/ vacuum chamber.

For reasons of size, weight and accessibility, the concept of an Instrument Mounting Platform (IMP) was adopted. Figure 3 shows schematically how all the parts of IRCAM are supported on this platform. The IMP is supported from ISU2 and the telescope rotator flange by eight struts and a braced bracket. With this arrangement the rotator can still be used and flexure of the IMP (which weighs 48 kgm) about the port is minimised.

IRCAM - the instrument itself, comprises two major parts, namely, a liquid cryogens container which is a modified Oxford Instruments LHe/LN$_2$ cryostat and a Camera Body/Vacuum Chamber made by Vacuum Generators. Within the Camera Body are several discrete modules, these are (i) the Filter Wheel Module and Lyot stop, (ii) the Camera Lens Module and (iii) the Detector Assembly. All of these components are mounted on an easily removable Optical Table which is supported from the outer Camera Body via low thermal conductivity (fibre glass) tubes. The table is cooled by flexible copper links to the liquid nitrogen cold shield.

There are 3 components within the camera head which must be driven from the outside. These are the two filter wheels and a fine focus drive for the detector. Each drive consists of identical 5-phase stepping motors, rotary vacuum feed-throughs and drive shafts with easily demountable couplings.

Cryogenics and detector temperature control

The need for a cryogenic instrument chamber stems from two sources both of which are fundamental to the ultimate, optimum use of the infrared array detector.

Firstly, the detector itself must be cooled in order to function as a photodiode, and the lower the temperature the lower its own thermal dark current, and therefore the higher its sensitivity. Secondly, the cooled detector consequently becomes very sensitive to the thermal (heat) radiation from all the warm components in the optical path and from scattered radiation outside the optical path, even if the detector is only sensitive out to a wavelength of 5μm. Thus the background signal on the detector can be very high unless as many as possible of the optical components themselves can be cooled (or are of low emissivity) and the detector protected by cold baffles.

Due mainly to the field-of-view requirements, the internal components of IRCAM are quite large. There is therefore a large thermal mass to be cooled and a large radiation loading on that mass due to the considerable surface area. For these reasons a large cooling capacity is required. After seriously considering closed-cycle coolers it was decided to remain with liquid cryogen containers for the time being until more information was available and some lab experience of such coolers had been obtained. Such work is currently under way. Liquid nitrogen temperatures are adequate for cooling the components within the Camera Head but the detector itself needs to operate somewhat colder than solid nitrogen - probably in the range 20-40K. For this reason a Liquid Helium (4K)/Liquid Nitrogen (77K) Cryostat was adopted and a scheme devised for establishing an intermediate temperature stage. The chosen cryostats have a 4 litre helium capacity and a 4.5 litre nitrogen capacity. The LN_2 reservoir is used to cool all the internal components and to sink the unavoidable heat inputs, the main component of which is radiation from the warm camera case. Similarly, cooling the detector to a temperature intermediate between LN_2 and LHe results in a heat input to the liquid helium. Typically, LN_2 hold times will be ~24 hours and LHe hold times ~36 hours. Since helium may not always be available on Mauna Kea and since IRCAM may be used in a "stand-by" (second instrument) mode or a "survey" mode in the background limit, it was essential to ensure that the inner vessel of the cryostat could be filled with liquid nitrogen and then attached to an evacuation pump to cool it to solid nitrogen temperature (50K). At the same time a method of "shorting" the detector assembly to this cold-face temperature was required. The solution adopted is that of a Variable Conductance Gas Switch (VCGS) as shown schematically in Figure 4 and described below.

The detector mounting block is supported off the optical table (which achieves ~85K) by two low conductivity fibre glass tubes. It is thermally connected by a high conductivity copper cold finger to the 'hot' end of a cylindrical chamber, called the Heat Exchanger, which is made of stainless steel (a poor conductor). The 'cold' end of this chamber is mounted on the copper cold face of the LHe vessel at 4K. The heat exchanger contains 4 concentrically separated copper cylinders, 2 hot and 2 cold. Connected to the exchanger by a stainless steel tube is a smaller copper chamber which is thermally linked to the cold

The SIRCAM package breaks down into three logically separate components. Firstly, there is a requirement to calculate the number of photons arriving above the telescope from astronomical objects of various kinds, from the thermal sky background and from night sky emission features. The former involves a knowledge of the typical anticipated brightnesses, space densities, spectral distributions and so on, for a variety of likely targets. The latter two items take into account thermal emission from the sky (using black body curves and empirically derived emissivities), and line emission from the upper atmosphere including absolute OH emission photon flux figures actually measured on Mauna Kea.

The second component of the package comprises a model of the effect upon these input photons due to the telescope optics, warm external optics, cold filters and cold re-imaging optics. This determines the final spectral and spatial distributions of source and background photons falling on the detector array, as a function of, for example, the scale-size employed. This naturally involves the convolution of the seeing disk with the PSF of the various optical components, and upon the reimaging scale employed for any particular observation.

The first two modules of SIRCAM essentially produce an "image" of input photon rates as measured by a perfect system just above the surface of the detector array. The actual detector array employed will, of course, not be perfect. The third component of the package attempts to model and simulate the response of the detector array to the input "image", from photons striking the array surface to the output signal voltage entering the preamplifier. Effects included in this section are, spectral quantum efficiency, readout noise, non-uniform response - including bad pixels, pixel filling factors, crosstalk and, in the case of photovoltaic DRC devices, the non-linear generation of dark (leakage) current and its overall effect on detector linearity.

Working through from the model astrophysical source of interest to the output image in this manner, we can examine the effect that variations of the great many parameters of a real infrared imaging system may have on the output "signal-to-noise", using this latter term in a general manner to describe the quality of, and rate at which, astronomy may be done. In addition, we are using SIRCAM (which is still in a state of continuous development) to stimulate scientific discussions about possible observing programs and to assist in the assessment of the system evaluation and commissioning tests as they are carried out.

The currently available test data on the SBRC 58 x 62 engineering-quality devices suggests dark currents of $<5 \times 10^{-4}$ pe.s^{-1}.pixel^{-1} at 50K, a quantum efficiency of ~80% at 3 μm and a readout noise of less than 500 electrons rms, which makes this detector a potentially useful device for many astronomical applications. Our simulations predict performances of the individual detectors equivalent to the best available single element detectors in the thermal background limited case, and at shorter wavelengths, if long on-chip integrations are possible, we should see significant improvements in sensitivity.

Conclusions

We have designed an infrared imaging system for the 3.8 m UKIRT which allows for several optical modes, and which employs 2-d solid state detectors of the type discussed elsewhere in these proceedings. In addition, we have developed software tools for the display, reduction, analysis and simulation of IR images. At present we are evaluating this system in the laboratory but we expect to move toward telescope trials later in the year.

Acknowledgements

The authors wish to thank their colleagues C. Aspin, R.J. Beetles, J. Clark, M.J. Paterson, I.A. Smith, J.M. Stewart and B.V. Vyas for their support and contributions to this project.

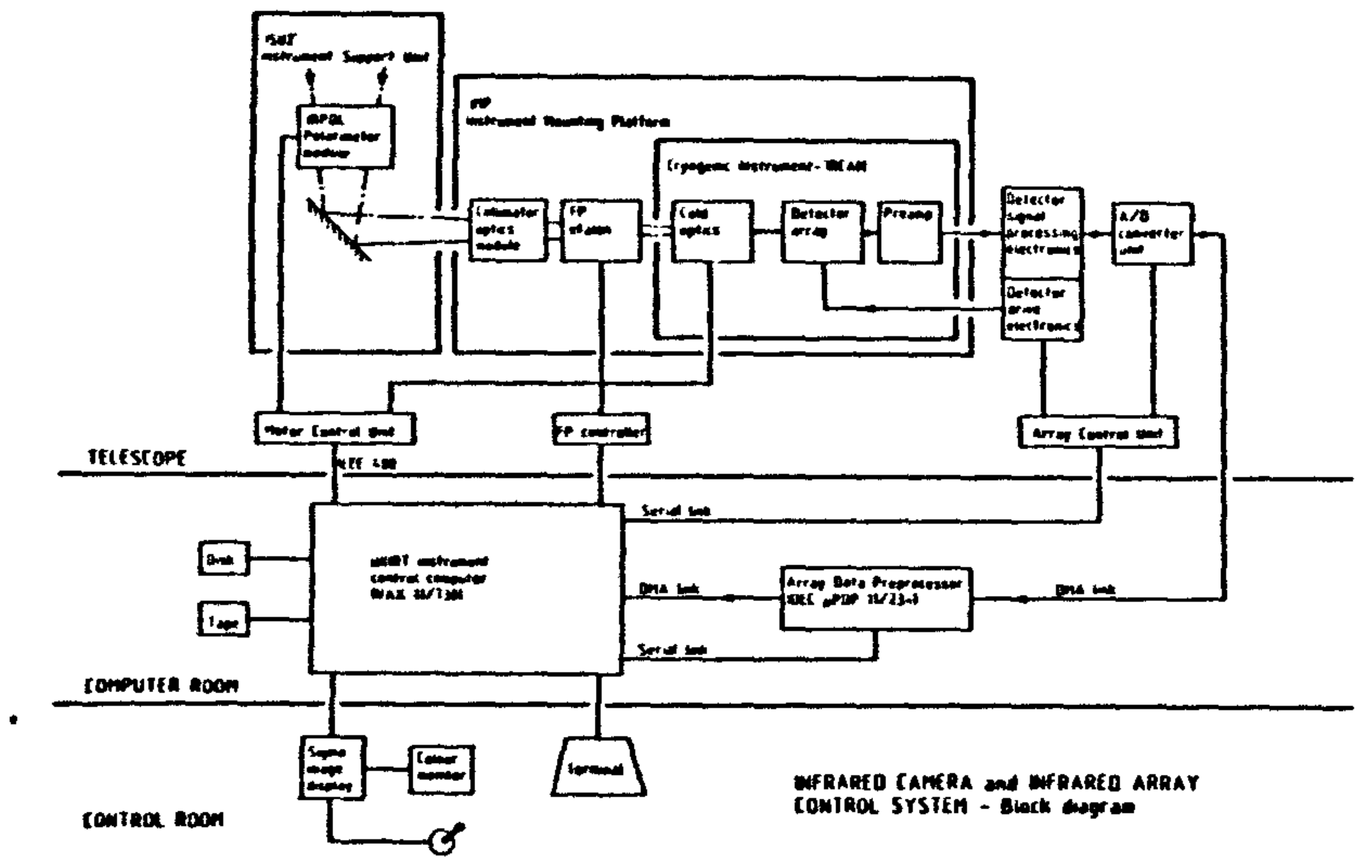


Figure 1. A block diagram of the UKIRT 1-5 μm Infrared Camera system.

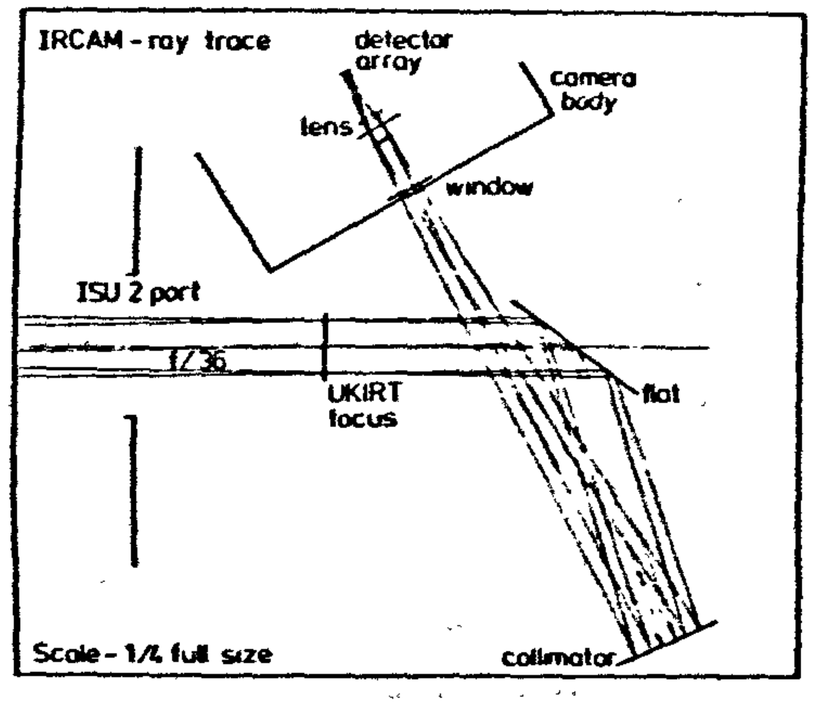


Figure 2. A plan view of the infrared camera optics with a ray trace.

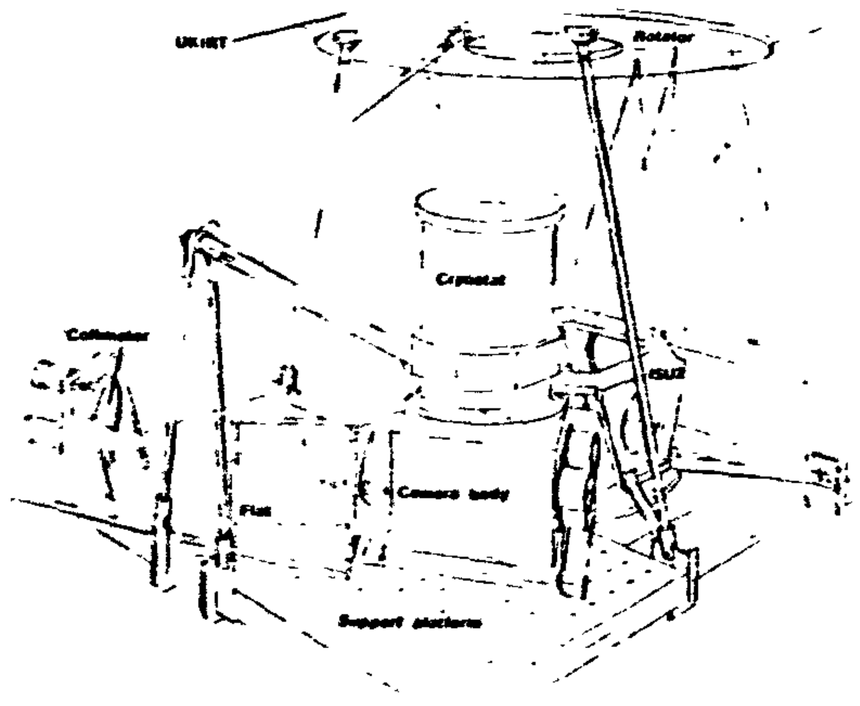
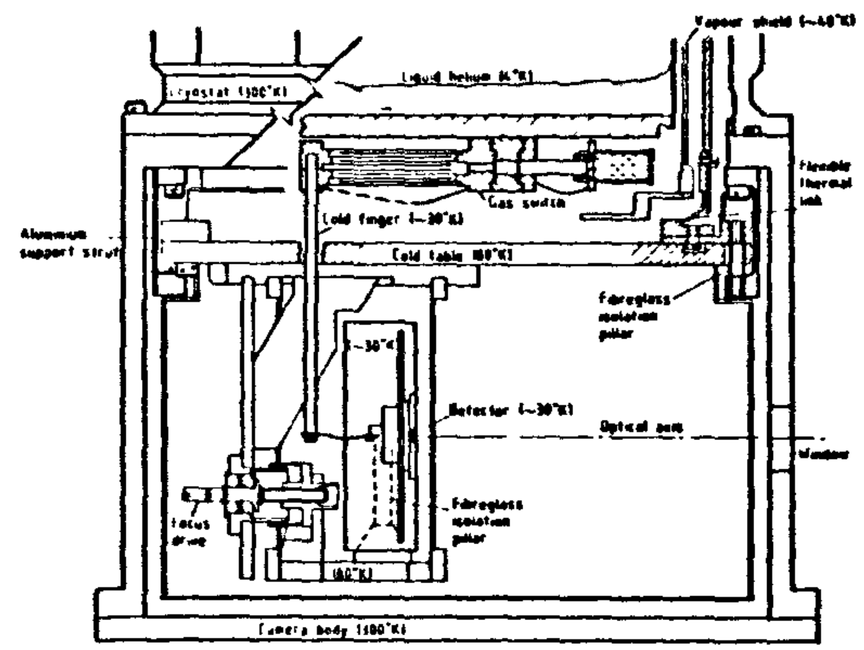


Figure 3. A schematic construction of how IRCAM will appear when attached to the 3.8m UKIRT.



CRYOGENIC THERMAL SUPPORT for the INFRARED CAMERA (IRCAM)

Figure 4. A schematic cross-section of the IRCAM camera body showing the principles of the thermal and mechanical design.

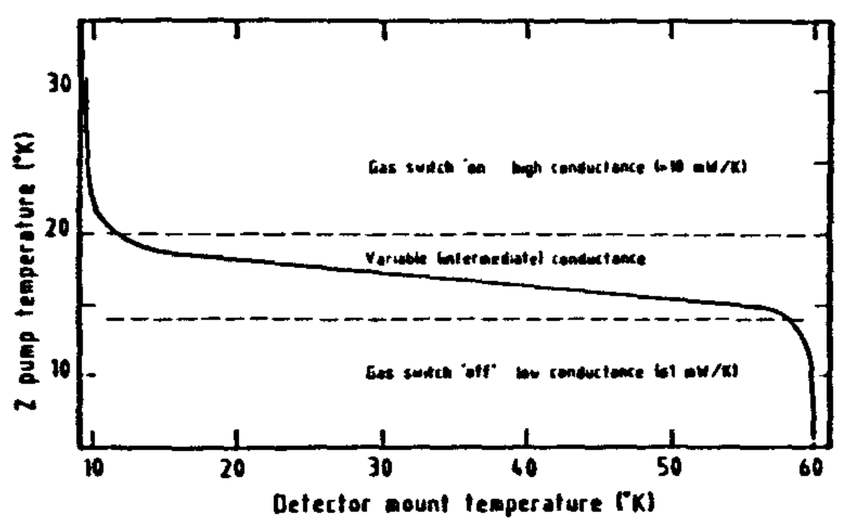


Figure 5. A typical characteristic curve for the gas switch showing the relationship between the temperature of the detector mount and that of the Zeolite sorption pump.

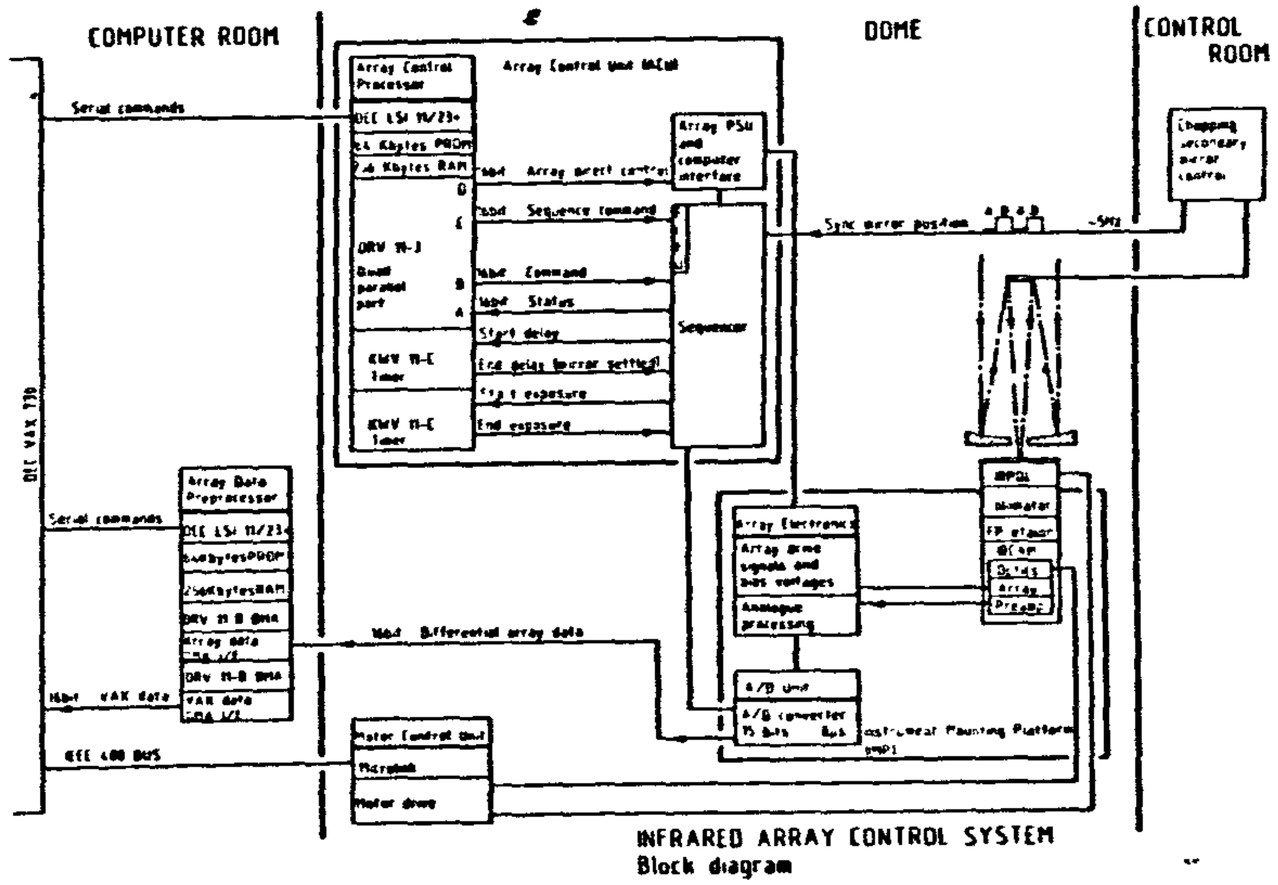


Figure 6. A block diagram of the IRCAM infrared array control and data acquisition system.

Mon. Not. R. astr. Soc. (1987) 225, 393–398

Discovery of an extended infrared reflection nebula around S106

I. S. McLean,[1] M. J. McCaughrean,[2] I. Gatley,[3] J. Hough,[4] S. Sato,[5] T. Nagata,[5] D. Axon,[6] M. G. Burton,[2] R. Garden,[2] C. Aspin,[1] T. Hasegawa,[7] M. Hayashi,[7] M. Morimoto[7] and N. Kaifu[7]

[1]*Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[2]*University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ*
[3]*UK Infrared Telescope Unit, 665 Komohana Street, Hilo, Hawaii 96720, USA*
[4]*Hatfield Polytechnic, PO Box 109, Hatfield, Herts*
[5]*Kyoto University, Kyoto 606, Japan*
[6]*Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire SK11 9DL*
[7]*Nobeyama Radio Observatory, Minamisaku, Nagano 384-13, Japan*

Accepted 1986 October 10. Received 1986 October 10; in original form 1986 July 18

Summary. An extensive, low surface brightness reflection nebula has been discovered around S106 by mapping the infrared polarization at 2.2 μm. Polarized flux is detected to almost 3 arcmin from the central source (IRS4), well beyond the main part of the optical H II nebula. The pattern of polarization vectors is centro-symmetric about the embedded object and is a characteristic of a reflection nebula. Polarizations as large as 25 per cent are observed, with the remarkable result that there is no obvious anomaly in the distribution of 2.2 μm polarization corresponding to the optical bipolar morphology. This difference can be explained by a simple model involving a compact dust disc near the embedded star. We also discuss the implications of this detection as an important new tracer of the extent of dust, and for the significant role that scattering from grains may play in protostellar clouds and obscured bipolar flows.

1 Introduction

S106 (OM1-99) is a bipolar nebula H II region which at optical wavelengths appears to have two luminous lobes separated by a well-defined dark lane. The source exciting the nebula has been identified at both optical and infrared wavelengths (Eiroa, Elsässer & Lahulla 1979; Sibille *et al.* 1975; Allen & Penston 1975; Pipher *et al.* 1976) and is embedded near the centre of the dark lane. We shall henceforth refer to the central source as IRS4, after Gehrz *et al.* (1982).

The nebula lies in a region about 20 arcmin in size which exhibits large-scale extinction and contains a massive molecular cloud (Lucas *et al.* 1978; Harvey *et al.* 1982; Kaifu & Suzuki-Hayashi 1986) with density maxima to the east and west of the compact H II region. Barnard's loop is located ~12 arcsec west of S106. Felli *et al.* (1984) and Dyson (1983) have reviewed the observational data in terms of a wind-driven bipolar model for S106. Furthermore, *IRAS* mapping data show this area to be just one small part of the Cygnus star-forming region which extends across many square degrees.

Polarimetric measurements of the brighter parts of the bipolar lobes of S106 established that IRS4 is the illuminating source (Lacasse *et al.* 1981; Tokunaga, Lebofsky & Rieke 1981; Perkins, King & Scarrott 1981; Staude *et al.* 1982; Gehrz *et al.* 1982). Recent high-resolution CCD observations by Aspin, McCaughrean & McLean (1986, to be published) from 0.5–0.8 μm have provided a more complete understanding of the scattering geometry of the inner regions. Almost all of the existing optical and infrared data are confined to a region smaller than 2 arcsec across and centred on IRS4, and infrared polarization observations exist for only the brightest parts of the lobes. In this paper we report extensive new observations in which we detect and map the distribution of polarized 2 μm flux at low surface brightness levels around S106. Such observations are a unique and powerful tracer of the influence of IRS4 and demonstrate that scattered light, and consequently dust particles of relatively large albedo, are present well out into the surrounding molecular cloud. The most spectacular discovery, however, is that the infrared polarization remains large along the E–W dark lane.

2 Observations and results

All of the new 2.2 μm polarization observations were made at the UK Infrared Telescope (operated on behalf of SERC by the Royal Observatory, Edinburgh) on Mauna Kea, Hawaii. Data were obtained on two separate observing runs namely, 1984 November 15–17 and 1985 July 23–26. The polarimeter module was provided by Kyoto University and the UKIRT photometer used was a single-channel InSb system (called UKT 9) with very low noise performance and with optics giving excellent beam profiles over large (~20 arcsec) apertures. For the 2.2 μm (K-band) observations a rotating single-order MgF_2 half-wave plate with a 50 mm clear aperture and 94 per cent transmission was used in the polarimeter module which in turn was located just above the dichroic beam-splitter in the UKIRT instrument support unit. A wire-grid analyser (0.25 μm spacing on a calcium fluoride substrate) was placed over the window of the side-looking UKT 9 dewar; the light loss at this polarizer was 1.0 mag.

Polarimetry was performed by sampling the photometer signal at eight points in the half-wave modulation cycle (approximately one point per second) and fitting a sine curve. For most of the observations an entrance aperture of 19.6 arcsec was used with an E–W chopper throw of almost 4 arcmin at a frequency of 3.5 Hz. For several positions in the nebula the reference beam was found to be contaminated (for one of the beam-switch positions) with polarized flux. These data were either not used or reduced as 'single-beam' if possible. During the second observing run a set of single-beam observations was carefully obtained to supplement the November run and, in addition, many points were repeated as a check. A few observations with a 12.6-arcsec aperture were also obtained for comparison.

Observations of the unpolarized star $\varkappa$ Ceti demonstrated that the instrumental polarization was less than 0.2 per cent, and polarization angles were corrected to the equatorial system to an accuracy of $\pm$ 1° by measurements of four highly polarized sources with known position angles (including GL 2591 and the BN object in Orion).

Fig. 1 shows the observed K-band polarization vector map superimposed on an R-band image (contour plot) of S106 to emphasize the extent of the 2 μm radiation. Each integration lasted 4 min

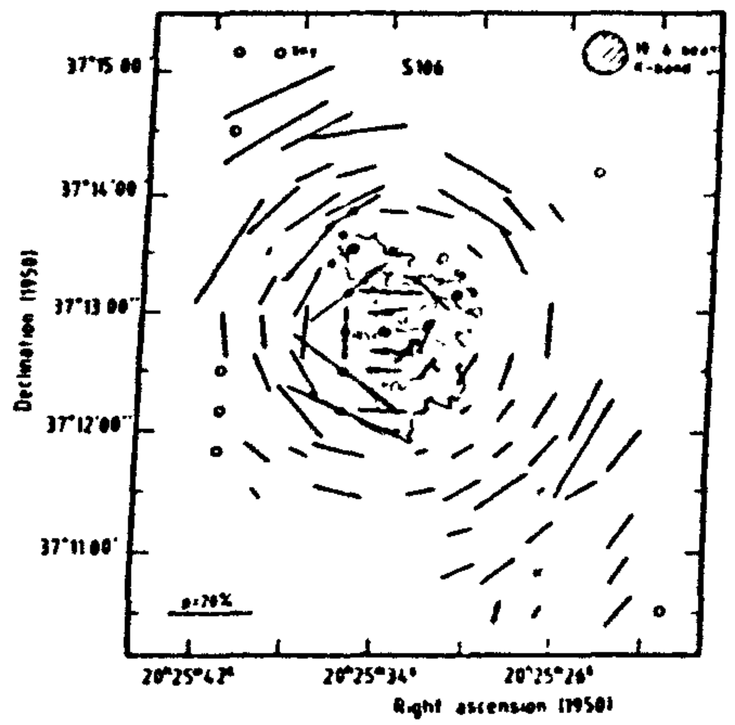


Figure 1. To form this composite diagram, linear polarization vectors measured at $K(2.2\,\mu m)$ have been overlaid on a contour plot of S106 taken at $R(0.75\,\mu m)$. The polarization measurements were obtained using the single-channel InSb photometer UKT 9 (with a broad-band K filter and a 19.6-arcsec aperture) and the Kyoto polarimeter on UKIRT. A filled circle indicates that a 12-arcsec aperture was used, an open circle indicates a flux level below our detection threshold and the cross indicates a point with definite polarization but with contamination. Double vectors represent repetitions on different observing runs.

(including beam-switching) and over 80 points were observed; several points were repeated and these are shown by a double line segment. The contour map of the optical nebula seen at $R(0.75\,\mu m)$ was obtained from the CCD data of Aspin *et al.* (1986, to be published). The lowest contour interval is ~17.0 mag per square arcsec and IRS4 is clearly detected. In Fig. 1 the size of the polarization vector (line segment) is proportional to the percentage polarization. Open circles indicate positions where the signal flux was indistinguishable from the sky level, and the cross marks a point with probable polarized flux but with contamination. Near IRS4 the large aperture (19 arcsec) was expected to cause problems due to the fact that the orientation of the vectors changes rapidly with position. A few 12.6-arcsec observations were also made, but even higher resolution would be preferable to study the innermost regions of the bipolar flow.

Polarizations as large as 25 per cent were observed to an extent of almost 3 arcmin, well beyond the detection limit of the optical H II region. To our knowledge this is the first time that a large detailed polarization map has been obtained at infrared wavelengths to such a low limit of surface brightness. Clearly, the polarization vector pattern is centro-symmetric about IRS4. Such a circular pattern of polarization vectors is expected for scattering of light into the line-of-sight by relatively small particles; the scattered radiation is polarized at right-angles to the scattering plane, that is orthogonal to the radius vector when projected on the sky. That the pattern is so remarkably centro-symmetric is very significant and implies that the illuminating source is very compact. The diagram also shows that the fractional polarization remains large and centro-symmetric in the region of the optical dark lane. If there is an extended disc-like obscuration due to dust in this region then it is optically thin at 2.2 μm. Close inspection reveals that the polarization is not uniform over the nebula; larger polarizations occur mainly to the east of the main N–S axis of symmetry of the H II region. Along some radius vectors the polarization

increases with angular distance in the classical manner of a reflection nebula while along others it drops off sharply. To the south-east and north-west, beyond 30 arcsec from IRS4, the surface brightness at 2 μm falls off to an average value of approximately 16.0 mag per square arcsec.

3 Discussion

The centro-symmetric polarization vector pattern is characteristic of simple scattering from a single illuminating source by particles which are small or of the same order as the wavelength. Since the nebular polarization of S106 is weaker at optical wavelengths than in the near-infrared (Aspin *et al.* 1986), this suggests that the particles are not *very* small and that second order scattering occurs at optical wavelengths or that intrinsic optical emission dilutes and washes out the optical polarization. Our optical data (Aspin *et al.* 1986, to be published) do not support the latter explanation. An interpretation in terms of a relatively optically thick dust cloud (at short λ) with dust grains of diameter ~0.1 μm would probably be sufficient to account for the observed wavelength dependence of polarization, but detailed grain models are needed.

In general, the net polarization along any given line-of-sight through a nebula depends on several factors, namely (i) the range of scattering angles encountered, (ii) the distribution (number density) of scatterers along the path length through the nebula and (iii) the amount of absorption before and after scattering. Large polarizations require that the depolarizing effects of line-of-sight integration be minimized, i.e. the range of scattering angles is intrinsically small or the distribution of scatterers and absorbers makes it appear so.

Since the total amount of scattered 2 μm flux is relatively large when integrated over the nebula, it is important to ascertain its source. On average, the surface brightness at 2 μm in the large bright annulus between 60 and 120 arcsec from IRS4 is $\sim 6\times10^{-24}$ W cm^{-2} (arcsec)$^{-2}$ and therefore the total scattered flux $F_{sca} \sim 2-10^{-12}$ W m^{-2}. It is straightforward to show that we cannot be observing scattered, 2 μm flux directly from the O9–B0 star ($T_* \sim 28000$ K, $R_* \sim 6$–$7\,R_\odot$ since, at the distance of S106 ($D=600$ pc) a K-band luminosity $>8\times10^{27}$ W is needed and even for a spherical geometry of perfectly reflecting dust grains the star yields only 3×10^{27} W at 2 μm. Clearly, one must attribute the source of 2 μm flux to the observed infrared object IRS4 which has reprocessed the short-wavelength radiation from the O9–B0 star. The observed colour temperature of IRS4 is 920 K (Gehrz *et al.* 1982) and the characteristic length scale for the source is presumably less than 0.1 arcsec, since no object larger than this is detected by the VLA (Felli *et al.* 1985). At a distance of 600 pc this angular size corresponds to $l \sim 10^{15}$ cm (~60 AU). Taking the radiating surface area to be $\sim\pi l^2$ the infrared luminosity is $\sim 7\times10^{29}$ W at 2 μm, over 200 times greater than the K-band flux from the star. The infrared source IRS4 is most likely to be either a disc of hot dust or associated with the stellar wind proposed by Felli *et al.* (1984).

The most significant feature of Fig. 1 is the fact that the 2 μm polarization remains large along the E–W extension of the optically dark line. In our view this optical extinction argues strongly for a dust disc and therefore it is not unreasonable to believe that IRS4 is the innermost part of this disc. Assuming then that the dark lane represents dust in the form of a large flattened region, it seems plausible that this dust is optically thin (to scattering) at 2 μm, i.e. the scattering optical depth $\tau_{sca} \leqslant 1$. According to Felli *et al.* (1984), the total extinction to the embedded star is possibly $A_V \sim 21$ mag or $A_K \sim 2$ mag and this implies that the total optical depth (τ) $\leqslant 2$ at 2 μm. Therefore the albedo (ω) of the dust grains (τ_{sca}/τ) is likely to be less than 0.5. In fact we can constrain the albedo further by estimating more carefully the amount of scattered flux toward the Earth [$F_{sca}(\lambda)$] and comparing this with the observed value. In the optically thin case

$$F_{sca}(\lambda) \simeq \omega(\lambda)\tau \frac{L_K}{4\pi D^2} \frac{\Omega}{4\pi}$$

where L_K is the predicted flux in the K-band from the infrared source, Ω is the solid angle subtended at the source by the part of the nebula being considered, and D is the distance to S106. To estimate L_K we use $L_K = L_* B(\lambda, T)\Delta\lambda/\sigma T^4$ where $B(\lambda, T)\Delta\lambda$ is the power emitted per unit area per steradian in the pass-band $\Delta\lambda$ at wavelength λ from a blackbody source at temperature T; L_* is the observed luminosity of the embedded star derived from far-infrared observations. With $L_* = 2\times10^4 L_\odot$, $D = 600$ pc, $T = 920$ K, $\lambda = 2.2\,\mu$m, $\Delta\lambda = 0.4\,\mu$m and $\Omega/4\pi = 0.5$ sr, then $F_{sca}(2.2\,\mu\text{m}) = \omega(\lambda)\tau \times 1.66\times10^{-11}$ W m^{-2}. Comparing this to our earlier observational estimate of $\sim 2\times10^{-12}$ W m^{-2}, we obtain $\omega\times\tau \simeq 0.12$. But since $\tau < 1$ then $\omega > 0.12$. Thus the albedo at 2 μm is estimated to lie in the interval $0.1 < \omega < 0.5$. The albedo of the dust grains around S106 therefore appears to be larger than expected (at 2 μm) for core-mantle grains ($\omega \sim 0.03$) and seems more consistent with silicates ($\omega \sim 0.14$); the addition of a component of water ice could raise the albedo further (Draine 1985; Rouan & Leger 1984). Infrared energy distributions of the knots embedded in the S106 dust clearly show the presence of silicates and that the 10 μm optical depth is small (Gehrz *et al.* 1982).

Adopting a mean grain size of $a = 0.15\,\mu$m, the mean column density for dust in the nebula (along the E–W lane) is $N_D = \tau/\pi a^2 Q_{ext} \leq 3\times10^{13}$ cm^{-2} for pure silicates (Draine 1985). Assuming that the dust uniformly fills a spherical volume ~ 0.6 pc in diameter, an upper limit to the mass of dust is $\sim 10\,M_\odot$. In reality only a fraction of the assumed volume is occupied by dust and gas. Evidence for cavities in this volume is the variation of polarized flux across the nebula and some correlation with the distribution of CS molecular gas (Kaifu & Suzuki-Hayashi 1986). There is also some evidence of symmetry, especially in the percentage polarization, about the 30° rotation axis suggested by the radio data (see, e.g. Eiroa *et al.* 1979).

It is also of interest to note that a direct (small aperture) measurement of the polarization of IRS4 at 2.2 μm yields $p = 1.5\pm0.2$ per cent at $56\pm5°$. This position angle agrees with the mean value derived (optically) for field stars near S106 (Lacasse *et al.* 1981). We suggest that this orientation probably specifies the axis of the *unresolved inner disc* (polarization vector perpendicular to disc); the observed polarization is caused mainly by scattering in the plane of the flattened disc. Relating the mean direction of the interstellar magnetic field to the orientation of the inner disc suggests that the disc was formed through cloud collapse along field lines (*cf.* Sato *et al.* 1985). In principle, the field orientation in the disc itself could be investigated by very high spatial resolution polarimetry.

Since the continuum polarization measurements prove that the influence of scattering can be traced to large angular distances from IRS4, it is quite possible that the surface brightness distribution of line emission may also be contaminated by scattering. This is of great significance for studies involving the mapping of bipolar outflows in molecular hydrogen lines to look for evidence of shocks or fluorescence. It is likely that the observed surface brightness distribution may not represent a physical distribution. We have recently found evidence of such effects in Orion and these will be reported elsewhere (Hough *et al.* 1987). It seems likely that such situations could be quite common among many similarly obscured compact H II regions but this can be tested in the future by polarization mapping in the $V=1\rightarrow0$ $S(1)$ line or the Q-branch of H_2. For S106 we have made one observation of the polarization of the 2.122 μm $S(1)$ line at the peak of the extended molecular hydrogen emission (17″ N, 5″ E of IRS4; Longmore *et al.* 1986). The result of an integration of 40 min gave $p = 7.2\pm1.8$ per cent at $\theta = 103\pm7°$, a 100 km s^{-1} Fabry–Perot etalon and the 2 μm CVF in UKT 9 were used to isolate the line. This result is not significantly different from the continuum value of 10 ± 1 per cent at $\theta = 110\pm3°$ thought to be due to scattering from IRS4, but a small component of unpolarized emission may still be present. Further measurements would be useful in locating the source of the H_2 emission.

In summary, if the presence of extended, low surface brightness nebulosity is a common feature of many heavily obscured infrared sources with bipolar flows, then polarization mapping at

2.2 μm with a photometer as sensitive as the UKIRT one becomes a powerful technique for tracing the angular size of the dust cloud. When the scatterers are related to the distribution of molecules and of shocked gas, one has a much more complete picture of the embryonic protostellar cloud. In principle, observations of scattered flux at other wavelengths give a way of probing the density (temperature) structure of the dust. Moreover, scattered radiation may seriously contaminate molecular line observations in some sources. In the case of S106, our new infrared polarization observations can be explained by scattering from dust grains of relatively high albedo, with the infrared source itself being interpreted as an inner disc of hot dust.

Acknowledgments

We are grateful to PATT for the allocation of telescope time, to Dolores Walther for her help at the telescope, and to Eric Becklin and Marcello Felli for helpful comments on this work. MJM, MGB and RG were supported on SERC studentships during this period. We also acknowledge support by the Japanese Ministry of Education.

References

Allen, D. A. & Penston, M. V., 1975 *Mon. Not. R. astr. Soc.*, 172, 245.
Draine, B. T., 1985. *Astrophys. J. Suppl.*, 57, 587.
Dyson, J. E., 1983. *Astr. Astrophys.*, 124, 77.
Eiroa, C., Elsässer, H. & Lahulla, J. F., 1979. *Astr. Astrophys.*, 74, 89.
Felli, M., Simon, M., Fischer, J. & Hamann, F., 1985. *Astr. Astrophys.*, 145, 305.
Felli, M., Staude, H. J., Reddmann, T., Massi, M., Eiroa, C., Hetele, H., Neckel, T. & Panagia, N., 1984. *Astr. Astrophys.*, 135, 261.
Gehrz, R. D., Grasdalen, G. L., Castelaz, M., Gullixson, C., Mozurkewich, D. & Hackwell, J. A., 1982. *Astrophys. J.*, 254, 550.
Harvey, P. M., Gatley, I., Thronson, H. A. & Werner, M. W., 1982. *Astrophys. J.*, 258, 568.
Hough, J. H. *et al.*, 1987. *Mon. Not. R. astr. Soc.*, in press.
Kaifu, N. & Suzuki-Hayashi, S., 1986. *IAU Symp. No. 115, Star Forming Regions*, preprint.
Lacasse, M. G., Boyce, D., Levreault, R., Pipher, J. L. & Sharpless, S., 1981. *Astr. Astrophys.*, 104, 57.
Longmore, A. J., Robson, E. I. & Jameson, R. F., 1986. *Mon. Not. R. astr. Soc.*, 221, 589.
Lucas, R., Le Sequeren, A. M., Kazes, J. & Encrenaz, P. J., 1978. *Astr. Astrophys.*, 66, 155.
Perkins, H. G., King, D. J. & Scarrott, S. M., 1981. *Mon. Not. R. astr. Soc.*, 196, 7p.
Pipher, J. L., Sharpless, S., Savedoff, M. P., Kerridge, S. J., Krassner, J., Schurmann, S., Soifer, B. T. & Merrill, K. M., 1976. *Astr. Astrophys.*, 51, 255.
Rouan, D. & Leger, A., 1984. *Astr. Astrophys.*, 132, L1.
Sato, S., Nagata, T., Nakajima, T., Nishida, M., Tanaka, M. & Yamashita, T., 1985. *Astrophys. J.*, 291, 708.
Sibille, F., Bergeat, J., Lunel, M. & Kandel, R., 1975. *Astr. Astrophys.*, 40, 441.
Staude, H. J., Lenzen, R., Dyck, H. M. & Schmidt, G. D., 1982. *Astrophys. J.*, 255, 95.
Tokunaga, A. T., Lebofsky, M. J. & Rieke, G. H., 1981. *Astr. Astrophys.*, 99, 108.

MAGNETIC FIELD STRUCTURE IN DARK CLOUDS

M. Tamura, T. Nagata, S. Sato, M. Tanaka
Department of Physics, University of Kyoto, Japan
N. Kaifu
Nobeyama Radio Observatory, University of Tokyo, Japan
J. Hough
Hatfield Polytechnic, UK
I. McLean, I. Gatley
United Kingdom Infrared Telescope Unit, UK
R. Garden, M. McCaughrean
Department of Astronomy, University of Edinburgh, UK

The magnetic field geometry in the central regions of two dark clouds has been mapped by measuring the polarization at 2.2 μm of background stars and of stars embedded in the clouds. The observations were done with the Kyoto polarimeter on the Agematsu 1m IR telescope in December 1984 for Heiles Cloud 2 in the Taurus dark cloud complex, and on the UKIRT 3.8m in May and July 1985 for the ρ Ophiuchus dark cloud core. The main results are:

i) Most of the stars in both regions show polarization and their maxima are 2.7% in Heiles Cloud 2 and 7.6% in ρ Oph, respectively. There are similar positive relations between polarization degree and extinction A_V's.

ii) The distribution of position angles for Heiles Cloud 2 shows a single mode at about 50° and that for ρ Oph shows a bimode, at about 50° and 150°.

iii) The magnetic fields, as delineated by the infrared polarization, appear perpendicular to the flattened elongations of the molecular clouds.

Although the efficiency of the alignment P_k/A_V in dense regions is lower than that in the vicinity of the clouds, it is still significant and would require a large magnetic field strength, if normal paramagnetic effects are required. The stratified structure of the clouds could be the result of a fragmentation process under the effect of the magnetic field.

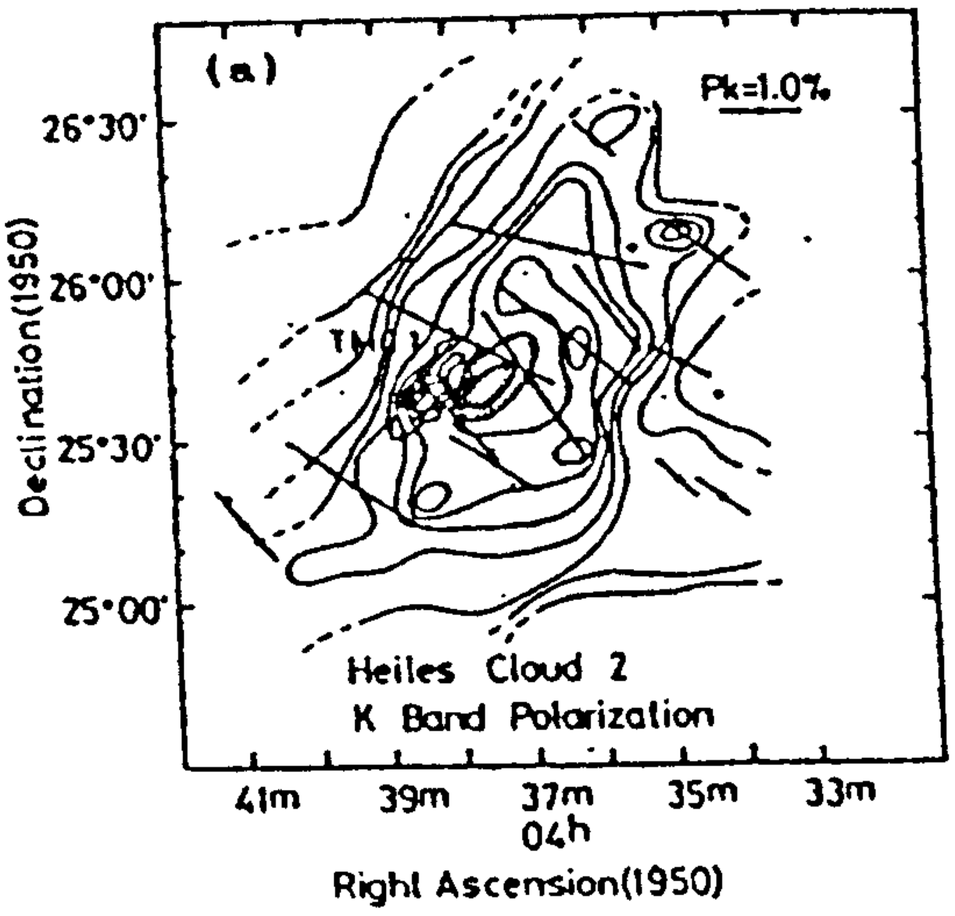
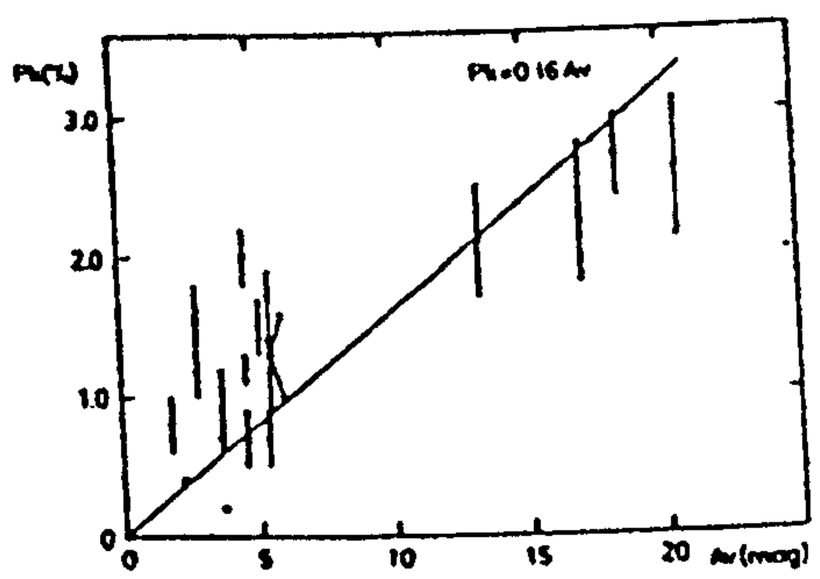


Figure 1

Fig. 1 (a) K-band polarization map towards Heiles Cloud 2 superposed on the A_V contour map (Sherwood and Wilson 1981). The CS emission contour map of TMCI (Snell, Langer, and Frerking 1982) is also shown. (b) K-band polarizations of the central core of the ρ Oph dark cloud are superposed on the map of the visual extinction (Wilking and Lada 1983).



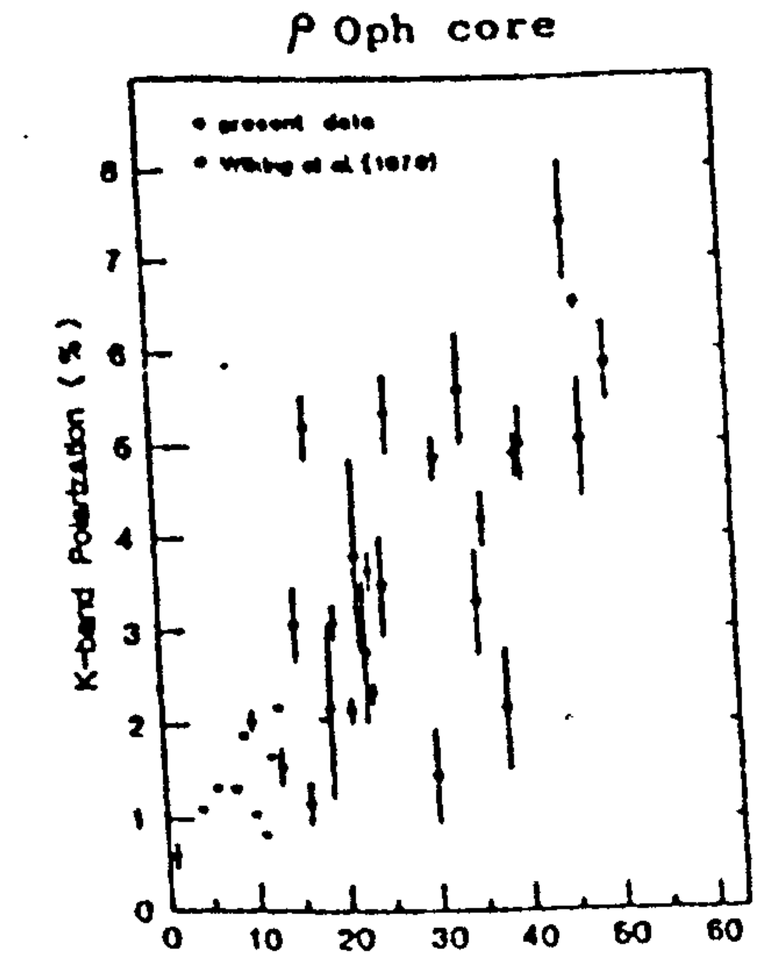


Figure 2

Fig. 2 Degree of polarization in the K-band versus total visual extinction A_V.

REFERENCES

Sherwood, W.A., and Wilson, T.L.: 1981, Astron. Astrophys. 101, 72.
Snell, R.L., Langer, W.D., Frerking, M.A.: 1982, Astrophys. J. 255, 149.
Wilking, B.A.,and Lada, C.J.: 1983, Astrophys. J. 274, 698.

MOUSCHOVIAS: What are the number densities in these clouds?
TAMURA: $\sim 10^3$ cm^{-3} in Heiles Cloud 2 and $\sim 10^4$ cm^{-3} in the ρ Oph core.
MOUSCHOVIAS: At such low densities ($\lesssim$ a few × 10 cm^{-3}), ambipolar diffusion cannot account for your derived, relatively small k(= 1/4) in the relation $(B/B_0) = (n/n_0)^k$. The uncertainties in deriving the field *strength*, as opposed to just its *direction*, from polatization observations must have something to do with it; aslo the values of B_0 and n_0 you use (*e.g.* see 1985, Astron. & Astrophys. 142, 41); they should be those values at which self gravity becomes important (and contraction perpendicular to field lines begins) and k becomes 1/2 (see 1976, Ap. J. 207, 141). The quantity B_0 can typically be 3 microgauss but n_0 is given by

$$n_0 = 137 \frac{B_0^{3/2}}{M^{1/2}} \text{ cm}^{-3} ,$$

where B_0 is the field measured in microgauss and M the mass measured in solar masses (see above reference, equation 5b).
PUDRITZ: Your observation of the alignment of the magnetic field of subcondensations with one another is extremely interesting. It suggests that the first structures to have formed in molecular clouds are large scale sheets, out of which subcondensations have fragmented. If subcondensations form first, one would see a more random orientation of their associated magnetic field orientations. Could you comment on this?
TAMURA: As I have shown in the talk, the polarization vectors are aligned fairly regularly; most of them are nearly perpendicular to the elongation of the clouds. One exception I think of is the "streamer" region to the east of the ρ Oph core. The polarizations there are parallel to the streams, but they are quite regular again.
FAZIO: Have you got any evidence of rotation or measurements of Doppler shifts across the structure for the same clouds where you observe magnetic fields polarized perpendicular to the elongation of the clouds?
TAMURA: No, I have no knowledge of rotation measurements for these complexes.
FAZIO: If not, it would be worthwhile to do those measurements, in order to investigate the role of magnetic field and rotation on the instability of these clouds (compare, i.e., with the Chandrasekhar and the Fermi Jeans mass).

SIMULATION AND MODELLING OF INFRARED CAMERA SYSTEMS

M. J. McCaughrean
Astronomy Department, University of Edinburgh
Royal Observatory, Blackford Hill,
Edinburgh EH9 3HJ, Scotland

I. S. McLean
United Kingdom Telescopes Headquarters
665, Komohana Street, Hilo, Hawaii 96720

Abstract. We describe a software modelling and simulation package (SIRCAM) implemented to aid the design and optimisation of new infrared instrumentation based around solid state 2-d detector arrays. We discuss the package with particular emphasis on results derived for a one to five micron imaging system (IRCAM) for the 3.8m UK InfraRed Telescope (UKIRT), based around an SBRC 62x58 InSb + SFD type array. The package is fully generalised however, and has also been used for other types of IR array instrumentation, including spectrometers, and SWIR/LWIR camera systems.

I. INTRODUCTION

The recent availability of arrays of detectors sensitive to infrared radiation has meant that true IR imaging is at last a reality. The spectacular gains obtained through the use of such arrays promise an exciting new period of discovery in the IR, as seen elsewhere in these proceedings. However, the design, implementation, and optimisation of instrumentation (both imaging and spectroscopic) employing IR arrays pose significant problems. Whilst there are lessons to be learnt from both current single channel IR instruments and solid state array detector instruments used in the optical, there are, in addition, novel problems associated with IR array instruments, not least the optimisation of the detectors themselves and the high data rates they can provide. Whilst building a first generation near-IR camera (IRCAM) for the UKIRT, we set about implementing a full software simulation (SIRCAM) of the camera system, with the aims and results described below. The package is implemented in Fortran-77, and is completely compatible with the ADAM environment, recently chosen as the standard data acquisition, reduction, and analysis environment for the UK's major observatory facilities around the world, and for the Starlink network of astronomical data processing computers.

II. AIMS

Our aims in developing the SIRCAM package were as follows :

(a) to provide feedback to the instrument design and operation
(b) to estimate system sensitivities across the wide range of possible observational configurations
(c) to explore the properties and limitations of particular detector arrays, especially with regards their photometric calibration
(d) to stimulate the development of data reduction and analysis software and techniques by providing realistically modelled data sets
(e) to provide a tool to help plan observations

III. INPUTS

SIRCAM comprises three main interfaced input sections, including the components listed below. All the inputs may be tailored by the user in order to explore various observational and instrumental configurations. At the same time, a range of default settings are available, and the user may save a particular configuration for re-examination at a later date.

(a) Astrophysics -
  (i) typical source flux and space densities
  (ii) broad band photometric magnitudes and colours for various classes of source
  (iii) simulated point source fields stored in catalogues
  (iv) simulated 'standard' extended emission sources
  (v) inter-unit conversions - magnitudes to janskys, janskys to watts per square centimetre etc..

The astrophysical inputs listed above are derived from standard texts and photometric catalogues, and used in order to provide realistic photon fluxes arriving above the atmosphere for a variety of astronomical sources. These photons are then 'processed' by the atmosphere, telescope, and optical train, accounting for the following effects :

(b) Atmosphere / Telescope / Optics -
  (i) transmission/emissivity of sky (IRTRANS model)
  (ii) non-thermal sky emission (e.g. OH)
  (iii) seeing profiles
  (iv) emission from telescope and warm optics
  (v) total system photon throughput
  (vi) image scalesize / spectrometer aperture
  (vii) cold filter profiles / spectrometer resolution
  (viii) optical point spread function

Finally, the 'processed' photons from both source and background arrive at the detector array. Considerable effort has been taken to model the behaviour of such arrays, including the following details :

(c) Detector Array -
- (i) dark current mechanisms (g-r, diffusion)
- (ii) quantum efficiency
- (iii) array uniformity (flat-field, dark current)
- (iv) non-linearities (photo-response, dark currents)
- (v) capacitances (junction, MOSFET)
- (vi) well capacity
- (vii) noise sources
- (viii) readout gain
- (ix) transfer efficiency
- (x) temperature dependencies of the above

IV. OUTPUTS

This package enables us to meet the requirements listed in Section II across the wide range of configurations applicable to IR array instrumentation, in the form of tables of numbers, graphical plots, and simulated imagery.

We have paid particular attention to the parameters relevant to the IRCAM design, namely that of an imaging camera operating at low (broad-band) and high (CVF, Fabry-Perot) spectral resolutions, at a variety of spatial resolutions (0.6 - 2.4"/pixel), on the 3.8m UKIRT, and with an SBRC 62x58 InSb + SFD type array. Two points of particular interest to us in this instance are the wide range of expected background flux rates, and the known non-linear dark current generation and photon response functions inherent in the SBRC architecture. Examples of SIRCAM output relevant to these topics are shown in Figures 1 and 2. Further exploration of the parameter space is precluded in these proceedings, but a full treatise on the inputs and outputs, the modelling algorithms, and results, are to be found elsewhere (McCaughrean 1987).

V. CONCLUSIONS

We have discussed the need for a software simulation package for new instrumentation based around IR array detectors. The SIRCAM package has been implemented in order to approach these questions, and it has been found to be of significant assistance in the design, testing, and astronomical commissioning of a new IR camera system for the UKIRT. The package is being further improved, and will be used alongside the development of new SWIR and LWIR cameras, and IR spectrometers.

Acknowledgments. We would like to thank Matt Mountain, Alan Hoffman, Mark Bird, Richard Wade, Terry Lee, Alan Tokunaga, Sandy Leggett, Clive Davenhall, Laurent Vigroux, and others, for valuable information and suggestions given during the development of SIRCAM.

References

McCaughrean, M. J. 1987, Ph.D. thesis, University of Edinburgh

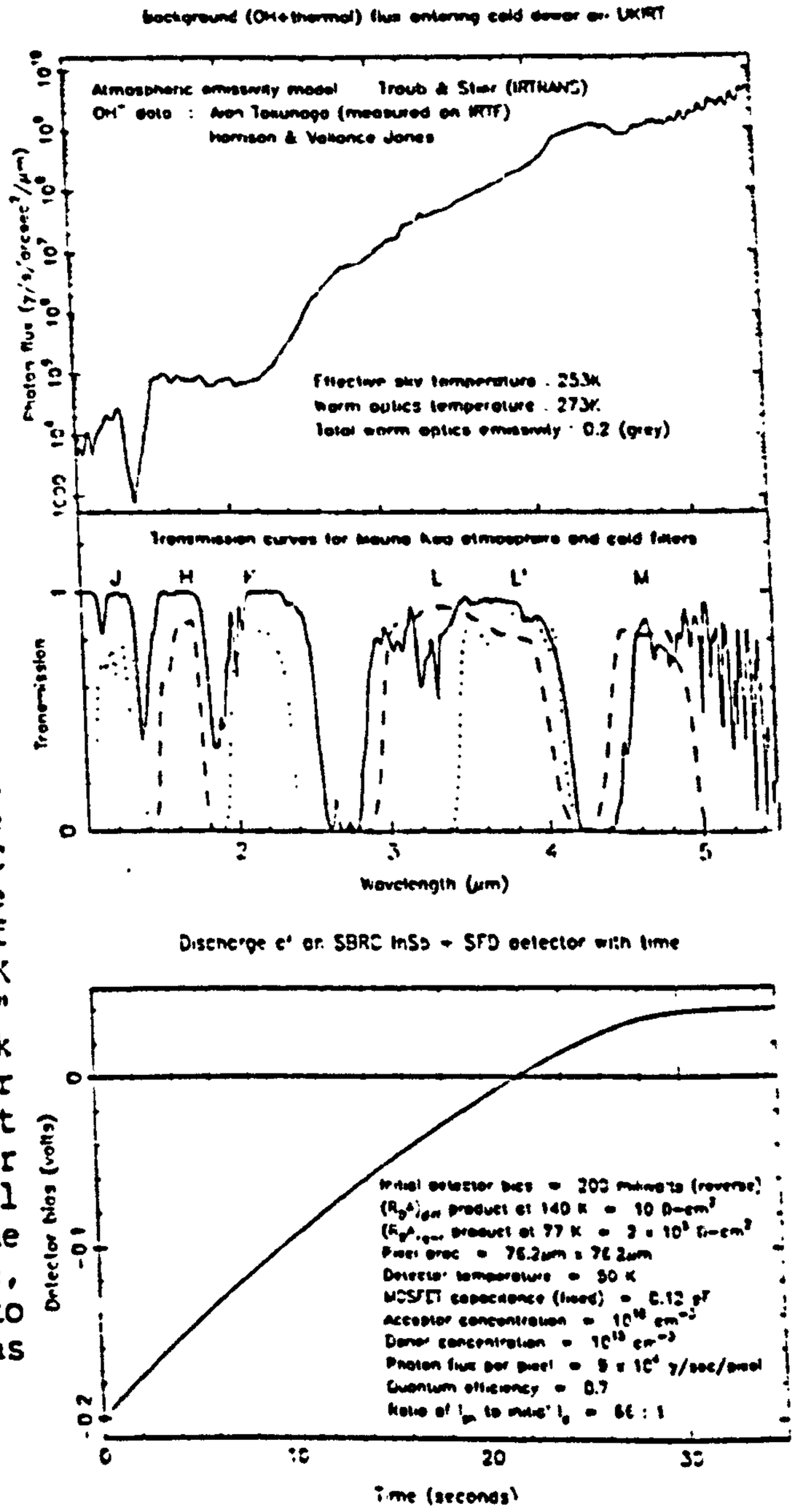


Figure 1. This shows the anticipated background arriving outside the window of a camera on the UKIRT at the summit of Mauna Kea. The background is due to thermal emission from the sky, telescope, and warm optics, along with short wavelength non-thermal OH emission. The cold optics will reduce this flux prior to reaching the detector. Also shown is the zenithal transmission for the sky above Mauna Kea, along with the transmission curves for the standard near-IR filter set.

Figure 2. Shown here is a simulation of the voltage discharge of an SBRC InSb + SFD detector. The photon flux is typical of that expected on UKIRT at K (2.2 microns) into 1.2" square pixels. The dark current at the detector temperature of 55K is just below 1,000 electrons per second per pixel at full reverse bias. Note the non-linear diode discharge, and that the detector runs to a steady-state forward bias after discharge.

DESIRED CHARACTERISTICS OF 2-D ARRAYS FOR SWIR IMAGERS AND SPECTROGRAPHS

R Wade, T J Lee, I S McLean and C M Mountain
Royal Observatory, Blackford Hill, Edinburgh

M J McCaughrean
University of Edinburgh, Department of Astronomy

I Baker
Mullard Ltd, Southampton

Abstract. High performance arrays are needed for the short wavelength infrared (SWIR) range, 1-2.5 microns, where requirements are rather special because of the low background in astronomy applications. The background as a function of wavelength for UKIRT on Mauna Kea is discussed and properties of existing arrays are used to derive desired characteristics for arrays for typical observations. These include high spatial resolution imaging and spectroscopy. It is shown that for imaging background limited performance can be obtained using existing CMT material if quieter multiplexers are successfully developed; however, high resolution spectroscopy challenges currently available material and multiplexers.

I Introduction

The ultimate performance limit on any infrared detector system will be that due to the shot noise on the background signal. For astronomy applications where the photon arrival rates from the sources of interest are generally very low and hours of telescope time per square metre of aperture are very expensive, a design goal of any system will be to be to achieve background limited performance. In the thermal IR this goal is normally easily achieved, however at the shorter IR wavelengths, 1-2.5 microns, the so-called SWIR region this has not usually be the case except for the very best of single channel systems. It is clearly important that we achieve background limited performance with future systems if we are to capitalize fully on the major investments being made in new facilities and instruments.

II The Background on Mauna Kea

The first step in determining the desired characteristics of arrays is to establish what background levels we expect at the telescope. In Figure 1, we show a plot of the expected rate of generation of photo-electrons for a typical instrument on a 3.8 metre telescope (McCaughrean 1987). A telescope emissivity of 20 per cent and total system throughput of 20 per cent have been adopted, with effective sky and telescope

optics temperatures of 253K and 275K respectively. While this plot will change for different atmospheric conditions and telescope temperatures it is representative of what one might expect on an infrared telescope on Mauna Kea. The curve clearly has two components. Longward of about 2.5 microns the background is dominated by thermal radiation from the telescope and atmosphere. Below 2.5 microns emission from OH in the atmosphere dominates.

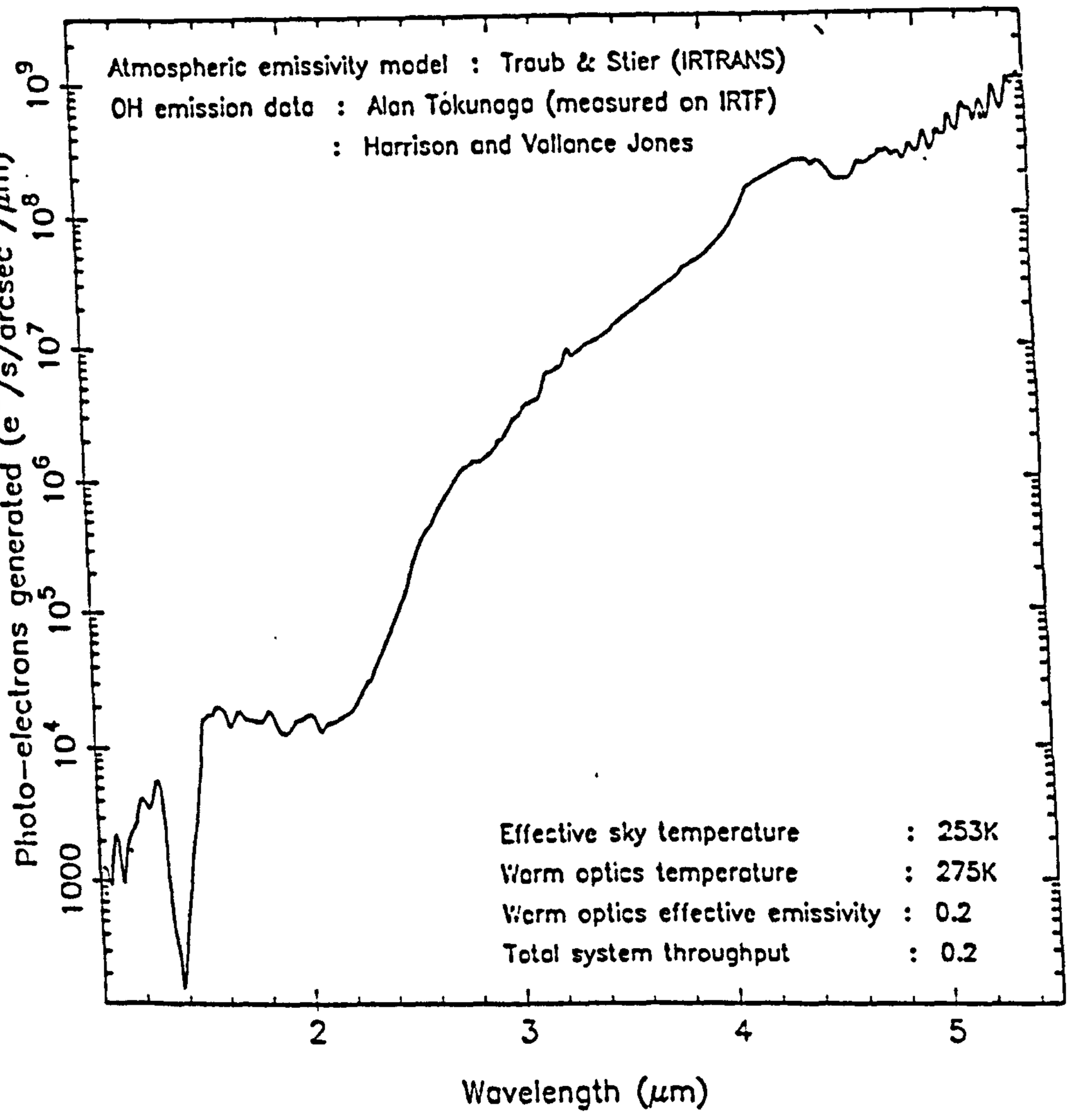


Fig. 1. The background on Mauna Kea. The number of photo electrons detected by a typical instrument on UKIRT is plotted against wavelength.

The backgrounds expected at 1 and 2 microns at various spectral resolutions of interest are plotted in Figure 2. We have assumed a 1 arcsecond square field and a 3.8m telescope. A feature of particular interest in this plot is the enormous range of expected backgrounds from less than 1 electron per second for high resolution spectroscopy at 1 micron to 10,000 electrons per second for imaging at 2 microns.

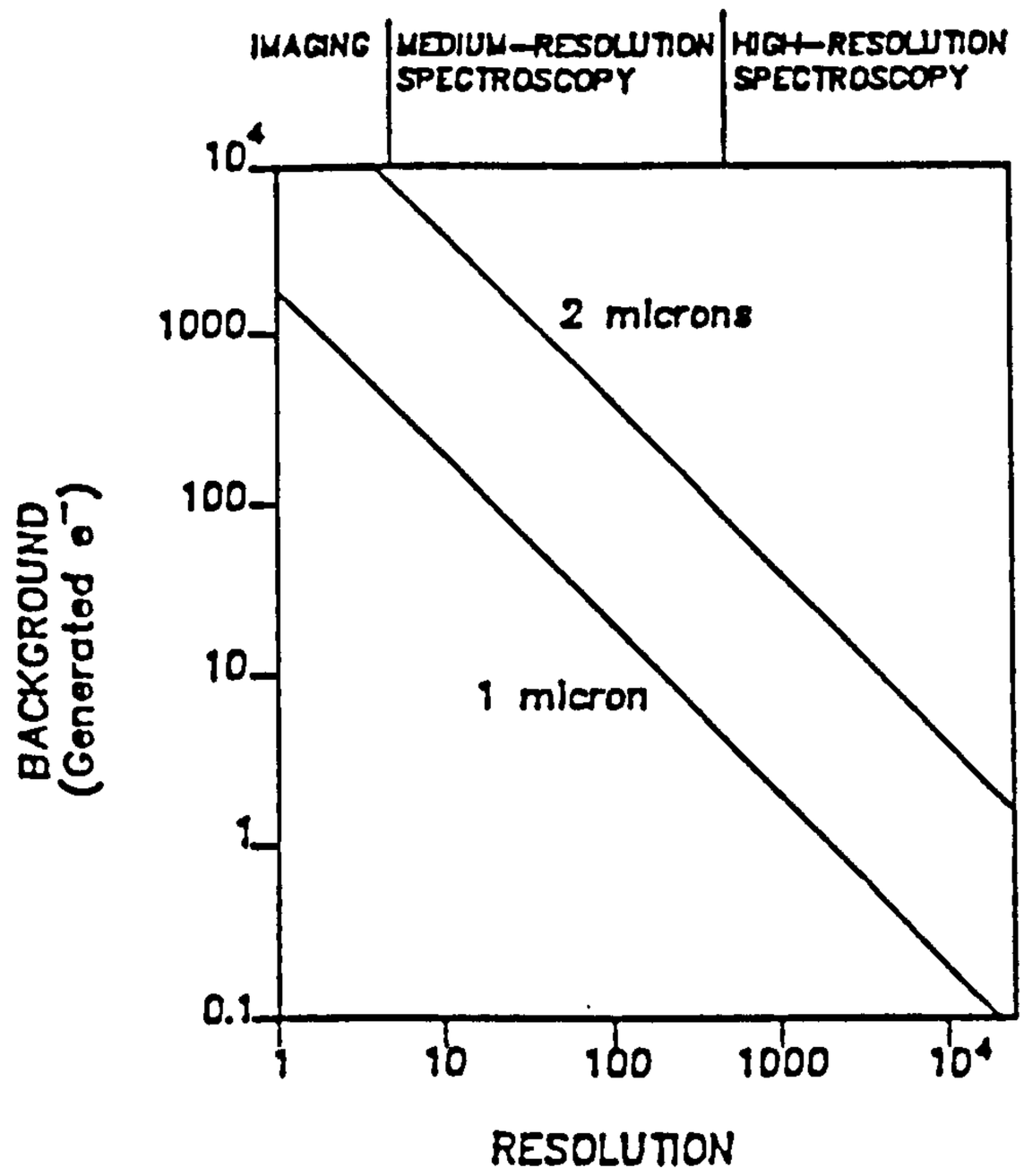


Fig. 2. Numbers of background generated electrons expected at 1 and 2 microns plotted against instrumental resolution.

III Requirements for Background Limited Operation

Given the expected backgrounds, what are the desired characteristics of a given detector array in order that it will be limited by this background? The first requirement is that the dark current in the detector N_{dk} is less than the background generated current N_b.

i.e. $$N_{dk} << N_b$$

This is clearly a necessary requirement if we are to avoid being limited by the noise on the dark current rather than that on the background current.

The second requirement is that we be able to integrate on-chip for a time T which is long enough that at the end of an integration the noise on the accumulated background signal will be greater than the read noise N_r

i.e. $$TN_b >> N_r^2$$

If this is not the case we will be limited by read noise rather than background noise. In order for this to be true there is a third requirement which is that the well capacity of the device is larger than the square of the read noise, this requirement does not appear to pose any problems for currently available devices.

i.e. $$\text{Well Capacity} >> N_r^2$$

The required minimum integration times to achieve background limited operation at 1 and 2 microns are plotted against spectral resolution in Figure 3.

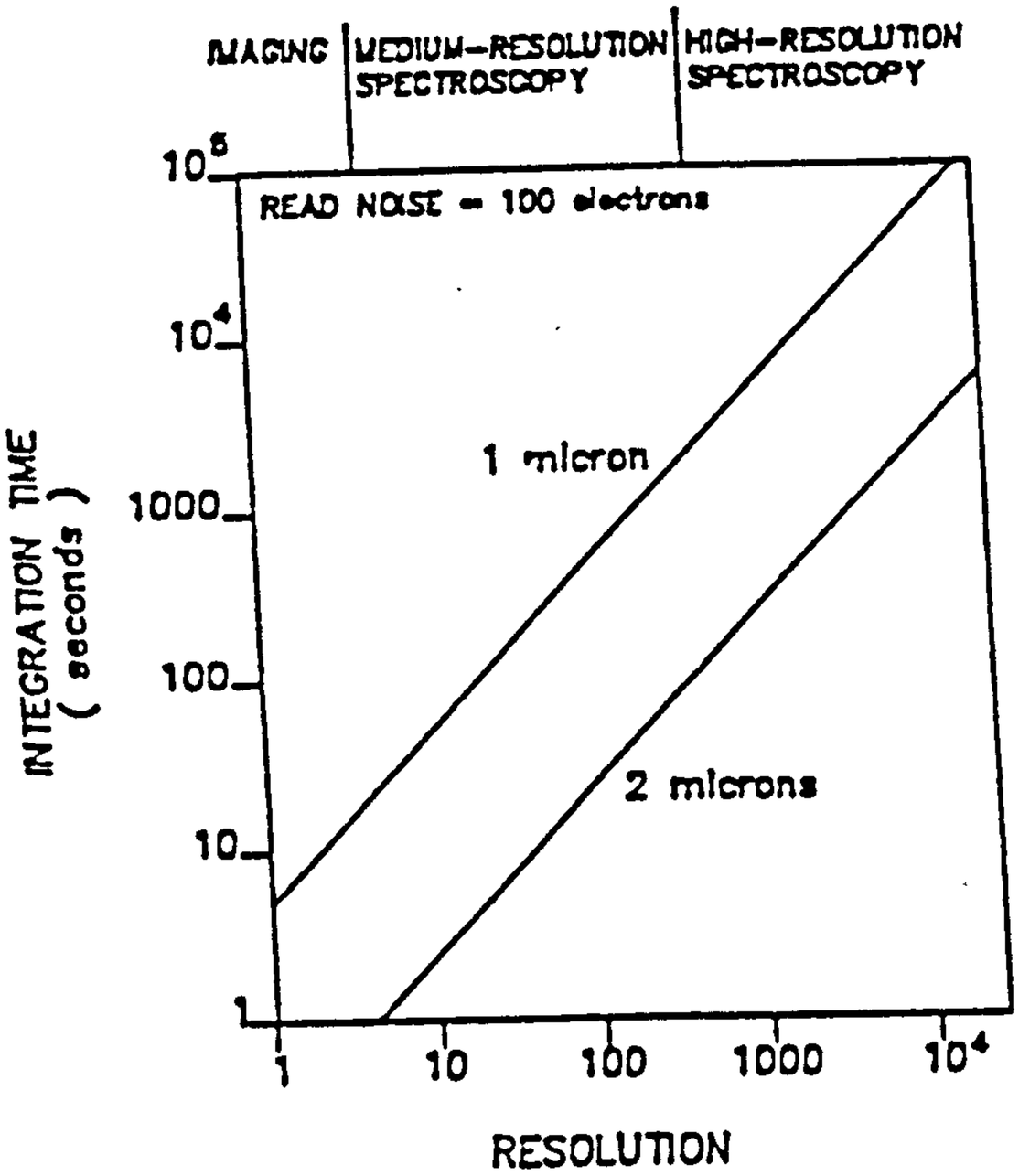


Fig.3. Required minimum integration times to reach the background limit at 1 and 2 microns plotted against instrumental resolution for a read noise of 100 electrons.

The important point to note in Figure 3. is that while for a read noise of 100 electrons, integration times of an hour or so will be sufficient to reach the background limit at 2 microns at all resolutions up to 10,000, they will not be sufficient for spectroscopy at 1 micron unless read noises can be reduced significantly.

IV Comparison of Required and Achieved Performances

The desired performance characteristics for arrays for imaging and spectroscopy in terms of dark current and read noise are summarised in Table 1.

TABLE 1

Desired characteristics to achieve background limited performance

	Imaging	Spectroscopy
N_{dk}	1000 - 10,000	< 1 - 10
N_r	100 - 500	50 - 100

Currently levels of achieved performance as reported at this meeting are summarised in Table 2. Performance figures are given for HgCdTe, InSb and Germanium diodes and are intended as a rough guide to performance rather than a definitive survey.

TABLE 2

Reported performances of currently available arrays

	HgCdTe	InSb	Ge
N_{dk}	10 - 1000	100 - 10,000	< 4
N_r	200 - 2000 reported 50 in prospect		

V Conclusions

Comparison of Tables 1 and 2 indicates that for imaging applications currently available HgCdTe devices should be able to achieve background limited performance in the 1-2.5 micron, SWIR, region. For spectroscopy applications both lower read noises and dark currents will be required. For the very shortest wavelengths germanium diodes may offer the very low dark currents required. For coverage of the whole 1-2.5 micron range however, high R_oA product HgCdTe material with a 2.5 micron cut-off wavelength operating at 50K looks promising. However, multiplexers with read noises of 50 electrons or less will be required.

REFERENCES

M J McCaughrean (1987). PhD Thesis, University of Edinburgh.

A.W. Harrison and A. Vallance Jones. Jnl.Atmos.Terr.Phys. 11, 192, 1957.

W.A. Traub and M.T. Stier. Appl. Optics 15, 364, 1976.

A. Tokunaga. (Private communication).